January 2009

Development and Implementation of Fully 3D Statistical Image Reconstruction Algorithms for Helical CT and Half-Ring PET Insert System

Daniel Keesing
Washington University in St. Louis

Recommended Citation

Keesing, Daniel, "Development and Implementation of Fully 3D Statistical Image Reconstruction Algorithms for Helical CT and Half-Ring PET Insert System" (2009). *All Theses and Dissertations (ETDs)*. 427.
https://openscholarship.wustl.edu/etd/427

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Biomedical Engineering

Dissertation Examination Committee:
Joseph A. O'Sullivan, Chair
Yuan-Chuan Tai, Co-Chair
Dennis L. Barbour
Daniel A. Low
Lihong Wang
Bruce R. Whiting

DEVELOPMENT AND IMPLEMENTATION OF FULLY 3D STATISTICAL IMAGE RECONSTRUCTION ALGORITHMS FOR HELICAL CT AND HALF-RING PET INSERT SYSTEM

by

Daniel Brian Keesing

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

August 2009

Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Development and Implementation of Fully 3D Statistical Image Reconstruction
Algorithms for Helical CT and Half-Ring PET Insert System
by
Daniel Brian Keesing
Doctor of Philosophy in Biomedical Engineering
Washington University in St. Louis, 2009
Research Advisors: Joseph A. O'Sullivan and Yuan-Chuan Tai

X-ray computed tomography (CT) and positron emission tomography (PET) have become widely used imaging modalities for screening, diagnosis, and image-guided treatment planning. Along with the increased clinical use are increased demands for high image quality with reduced ionizing radiation dose to the patient. Despite their significantly high computational cost, statistical iterative reconstruction algorithms are known to reconstruct high-quality images from noisy tomographic datasets. The overall goal of this work is to design statistical reconstruction software for clinical x-ray CT scanners, and for a novel PET system that utilizes high-resolution detectors within the field of view of a whole-body PET scanner. The complex choices involved in the development and implementation of image reconstruction algorithms are fundamentally linked to the ways in which the data is acquired, and they require detailed knowledge of the various sources of signal degradation. Both of the imaging modalities investigated in this work have their own set of challenges. However, by utilizing an underlying statistical model for the measured data, we are able to use a common framework for this class of tomographic problems.

We first present the details of a new fully 3D regularized statistical reconstruction algorithm for multislice helical CT. To reduce the computation time, the algorithm was carefully parallelized by identifying and taking advantage of the specific symmetry found in helical CT. Some basic image quality measures were evaluated using measured phantom and clinical datasets, and they indicate that our algorithm achieves comparable or superior performance over the fast analytical methods considered in this work. Next, we present our fully 3D reconstruction efforts for a high-resolution half-ring PET insert. We found that this unusual geometry requires extensive redevelopment of existing reconstruction methods in PET. We redesigned the major components of the data modeling process and incorporated them into our reconstruction algorithms. The algorithms were tested using simulated Monte Carlo data and phantom data acquired by a PET insert prototype system. Overall, we have developed new, computationally efficient methods to perform fully 3D statistical reconstructions on clinically-sized datasets.

Acknowledgments

It has been a privilege working with my advisors, Jody O'Sullivan and Yuan-Chuan Tai, on their projects for these past several years. I have learned an incredible amount from them and their research groups. I am grateful for the trust they had in me to do much of my work independently. Both of them have always been eager to support me and provide me with excellent opportunities. I would like to thank my thesis committee for taking time out of their busy schedules to see me through this process. Kurt Thoroughman was my academic advisor in Biomedical Engineering.

I would also like to acknowledge my colleagues and labmates (past and present) who have helped make this research possible, namely, Dave Politte, Bruce Whiting, Mike Harrod, Debashish Pal, Geoff San Antonio, Liangjun Xie, Norbert Agbeko, Heyu Wu, Sergey Komarov, and Tae Yong Song. In particular, I am fortunate to have worked closely with Debashish and to have taken part in many interesting discussions with him. John Young and Stefan Siegel at Siemens Medical Solutions in Knoxville have also been of great assistance regarding the half-ring insert PET system.

I would most definitely be in a different area of research (or perhaps not even a graduate student at all) if I had not met Boris Hasselblatt and Misha Kilmer at Tufts University. I had done undergraduate research with both of them, and shortly discovered that I had a passion to work in the field of medical imaging.

My parents and brother Jeff have provided me with unlimited love and support, even through tough times. They would have preferred to see me more often, but understood the commitments involved in pursuing graduate-level research. The rest of my family, but especially my aunt, have been extremely supportive as well. I have also met some wonderful friends at Wash U, and am very fortunate to be so close to them. Dick Wu in particular has been a truly remarkable friend and housemate these past five years.

Finally, I would like to acknowledge my financial support from the National Science Foundation Graduate Research Fellowship Program. This work was also supported in part by NIH grant 5R01CA075371-08 (J. F. Williamson, PI), Susan G. Komen for the Cure grant BCTR0601279 (Y.-C. Tai, PI), and NIH grant 5R33CA110011-04

(Y.-C. Tai, PI). Support was additionally provided in part by the National Science Foundation through TeraGrid computational resources at the National Center for Supercomputing Applications (University of Illinois at Urbana-Champaign).

Daniel Brian Keesing

Washington University in Saint Louis
August 2009

Dedicated to my family, for their love and support every step of the way.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Medical imaging plays a vital role in the diagnosis of numerous diseases and injuries, and is also critical in a growing number of image-guided therapy applications. It encompasses a wide array of technologies that collectively result in image resolutions ranging from picometer (e.g., electron microscopy) to millimeter (e.g., clinical imaging modalities).

Images are usually categorized as being either functional or structural, depending on the contrast mechanism being imaged. For example, in-vivo optical imaging techniques may measure the absorption of deoxygenated and oxygenated hemoglobin in the blood (functional), while changes in x-ray attenuation may be used to distinguish between soft tissue and bone (structural). There has recently been a push towards multi-modality imaging, in which registered functional and structural images are acquired at nearly the same time to enhance the value of having either image alone.

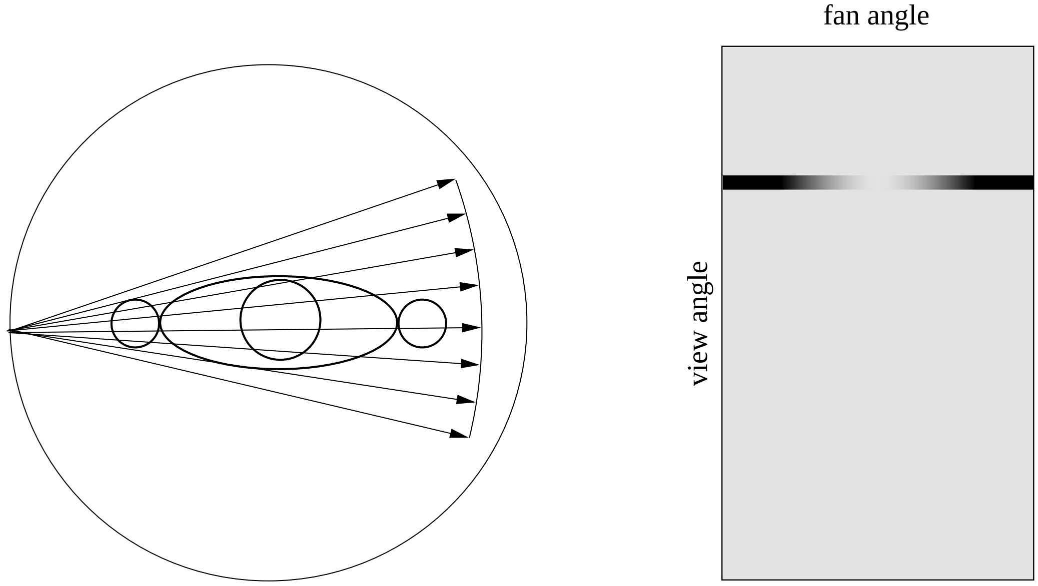
In general, an imaging system is designed to measure some spatially-variant (and possibly time-variant) contrast mechanism given a specific type of energy input. The image formation process requires a priori knowledge of how the incident energy physically interacts with biological tissue and how it is then detected by the system. This energy could be in the form of mechanical pressure waves, as is used in ultrasound imaging, or it could be in the form of electromagnetic radiation, such as visible light, infrared, radio waves, x-rays, and gamma rays. To carry out an imaging task, at least one radiation source and an array of detectors are needed. The detectors are optimized to detect the form of energy that results from interactions with the body. In the case of a digital imaging system, the detector array must also be able to convert its measurements at each channel into electrical signals that can be transferred to and received by a computer. Finally, after any necessary data or image processing, the image can be displayed for visual interpretation, or can be used in the extraction of other quantitative data.

Unlike medical imaging modalities that acquire images directly, like optical microscopy, we will concern ourselves here with tomographic imaging systems, in which some known *function* of the image is acquired through indirect measurements. As will be discussed in great detail in this dissertation, some form of image reconstruction method is needed after tomographic data acquisition to obtain an actual image. The fundamental significance of the image reconstruction process is that it enables one to generate volumetric images of the body's interior from the non-invasive measurements taken outside the body.

The location of the radiation source and detectors relative to the body determines whether the modality is considered transmission or emission tomography, or some other type. In transmission computed tomography (CT), an external source transmits photons of energy through the body. Some of these photons will be detected on the other side of the body, depending on the attenuation characteristics of the tissue being imaged. In single-photon emission computed tomography (SPECT) and positron emission tomography (PET), the source is first allowed to distribute itself to target molecules inside the body through the bloodstream or lungs. Its emitted photons of energy are then detected outside the body.

CT, SPECT, and PET all depend on the fact that their respective source photons travel along relatively straight paths through the body. At the other extreme, diffuse optical tomography systems are designed to detect the source photons (near-infrared light) after having scattered many times in the tissue; as such, they can be configured in either a transmission or reflection geometry (depending on the application).

The next two sections will highlight the basics of the CT and PET imaging modalities, followed by a third section describing a novel PET scanner geometry called virtual pinhole PET. Collectively, these modalities will be the main focus of this dissertation.



(a) (b)

Figure 1.1: Cross-section of a CT scanner. (a) The x-ray source travels around the circle in conjunction with an opposing arc of detectors on the other side of the patient; (b) sinogram array, where each row corresponds to a fan-beam projection at a particular view angle.

1.1 CT Data Acquisition

CT imaging uses a collimated x-ray source and an opposing array of detectors to capture projections from many angles, called views, around the patient. Other radiation sources may be used in place of x-rays, such as gamma rays (for the purpose of attenuation correction in PET imaging). Figure 1.1(a) shows a third-generation CT scanner in the process of acquiring fanbeam data in a single slice. For each slice to be imaged, the gantry rotates at most 360° around the patient. At each view angle, a large number of photons are randomly generated by the x-ray tube and collimated to the shape of the fanbeam defined by the detector arc. As x-rays pass through the body, they are either absorbed (by the photoelectric effect), scattered (by Compton scattering), or transmitted in a straight path (without interaction). The spatially-variant linear attenuation coefficient distribution of the tissue determines which parts of the incident x-ray beam are attenuated, and by how much. Detector elements in the opposing array measure the resulting attenuated x-ray beam, at which time the scanner transfers the result to a computer for storage and subsequent reconstruction. Specifically, the detector measurements at a particular view angle are placed into their respective row in a sinogram, as shown in Figure 1.1(b).

The primary means of attenuation in tissue at diagnostic x-ray CT energies is by photoelectric absorption, in which all of an x-ray photon's energy is absorbed by an electron. When a photoelectric interaction occurs, the x-ray imparts sufficient energy on the electron for it to overcome its binding potential, thereby ejecting it from its atomic orbit. This aspect of x-ray CT is undesirable, since the ejection of an electron has the effect of ionizing the atom. The ionizing radiation dose to a patient from a CT scan has been raising concerns in recent years about the potentially increased risk of developing cancer [43]. However, the benefit of the information obtained by CT scans often greatly outweighs this risk.

Although single-slice scanners (like the one shown in Figure 1.1) are now obsolete, their geometry still provides the underlying basis for a large class of reconstruction algorithms. Current state-of-the-art scanners utilize cone-beam measurements, which have a wider axial coverage (i.e., more than one detector row in the axial direction). This allows a larger field of view (FOV) to be scanned in the same amount of time that it would take to scan a single slice. Although the cone-beam geometry complicates scatter rejection and makes it challenging to obtain enough samples in the outer slices of the projection, it significantly speeds up the acquisition, thereby making applications like cardiac imaging possible.

Scanners with a larger axial extent are not restricted to a circular orbit; helical source trajectories are also widely used in practice, in which the patient bed translates through the scanner bore while the gantry rotates simultaneously. The remarkably fast gantry rotation speed of current helical scanners ($\sim$ 0.3 seconds per rotation) enables a whole body scan to be completed in a matter of seconds. The fast scan time also makes helical CT extremely useful for cardiac and pulmonary exams, rapid patient evaluation in an emergency room setting, and many other applications.

1.2 PET Data Acquisition

Unlike CT, PET is a functional imaging modality. It is used as a powerful molecular imaging tool for clinical and pre-clinical applications in oncology, cardiology, and neurology. PET imaging requires the uptake of a radiotracer by the body, which indirectly behaves as an internal radiation source. The radiotracer consists of a radioactive isotope that has been chemically attached to, or incorporated into, some pharmacologically relevant molecule.

The most commonly used radiotracer in current clinical practice is fluorine-18 fluorodeoxyglucose (^{18}F-FDG). FDG is a glucose analog, meaning it is taken up by the body in the same way that glucose is. By using the positron-emitting fluorine isotope ^{18}F in the FDG molecule, radioactivity will become concentrated in cells with high glucose uptake. As glucose is metabolized by a cell, it becomes phosphorylated, thereby preventing it from readily leaving the cell. Patients who undergo a PET scan with ^{18}F-FDG typically must have fasted for several hours pre-injection, and wait about an hour post-injection, for the radiotracer to adequately distribute itself in the body. Malignant tumors are often the most metabolically active cells in the body, and will therefore take up the largest proportion of ^{18}F-FDG. Over time, the ^{18}F will also be found in organs responsible for clearance, such as the kidneys and bladder.

Each radioactive atom in a PET radiotracer will spontaneously emit positrons over time according to the Poisson random distribution (with mean value equal to its activity, i.e., emission rate). Each emitted positron will move in a random walk around the nearby tissue (an effect known as positron range) before ultimately annihilating with an electron. As shown in Figure 1.2, the annihilation of the positron and electron results in the production of two gamma rays that travel approximately 180° apart from each other. If the positron was not at rest when the annihilation took place, the momentum is conserved by the gamma ray pair in the form of some small angular deviation from 180°. This undesirable effect is known as gamma ray acolinearity.

A typical PET scanner with patient lying in the FOV is shown in Figure 1.3. The detectors fully encircle the patient within a small axial region to achieve maximum sensitivity. “Prompt” coincidences are recorded by the scanner if two gamma rays are detected within a tight coincidence timing window. Each recorded event is specified

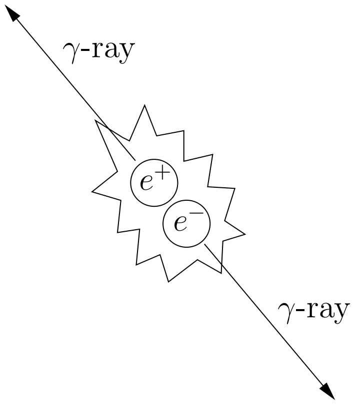


Figure 1.2: Positron emission leads to annihilation with a nearby electron, producing two gamma rays that are approximately 180° apart.

by the pair of detectors involved in the coincidence detection; this pair defines what is called a line of response (LOR). The key point of PET imaging is as follows: a coincidence detection implies that a positron source was located somewhere along the LOR connecting the two detectors that registered the event. In practice, however, each LOR not only contains the true coincidences, but also randoms and scattered counts. Randoms are due to the near-simultaneous detection of two gamma rays from independent annihilation events, and scattered counts are due to one or both gamma rays having scattered in the object or detectors. Randoms and scatter degrade image contrast if not properly accounted for in the reconstruction procedure since they constitute additive background. Additionally, positron range, gamma ray acolinearity, and several other effects degrade the spatial resolution of the measured data.

PET holds great promise for future breakthroughs, as a properly designed radiotracer can in principle target specific types of molecular receptors on cells. ^{18}F-FDG is not specific to cancer cells, but is generally taken up by these cells in higher quantities than normal cells. As our understanding of molecular and cellular biology expands, novel radiotracers can be developed that will provide ever greater specificity to certain diseases. Higher specificity helps improve the accuracy of PET imaging for diagnostic and treatment monitoring purposes, and provides pharmaceutical companies with the ability to evaluate drug efficacy (e.g., pharmacokinetics) more quantitatively.

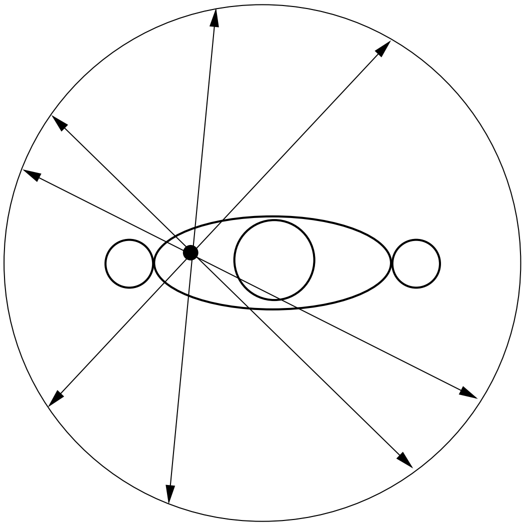


Figure 1.3: Cross-section of a PET scanner. A positron source is located at the black dot within the patient, and its indirectly-produced gamma ray emissions are detected in coincidence by pairs of detectors located all around the circle. In the ideal PET imaging scenario, a positron source physically resides somewhere along each line in which a coincidence event was detected.

1.3 Virtual Pinhole PET

There are several fundamental factors that limit the spatial resolution of a conventional circular PET scanner. The most basic factor is the intrinsic resolution of the scintillating crystals, which is determined primarily by the crystal size. However, even if high-resolution detectors are used throughout the scanner, gamma ray acolinearity degrades the resolution proportionally to the distance D between two detectors. The full-width at half-maximum (FWHM) blurring caused by the angular deviation from 180° annihilation is normally written as $R_{\text{aco}} \approx 0.0022 \cdot D$. For a 90 cm diameter scanner, this corresponds to ~ 2 mm FWHM blurring at the isocenter. Another factor responsible for resolution degradation in PET imaging is positron range. The severity of this effect depends on the radioisotope; ^{18}F has sub-millimeter positron range in biological tissue, which is negligible in comparison to the aforementioned factors. There are many other reasons for degraded system resolution, such as intercrystal penetration and intercrystal scattering within a block of crystals, and the photo-multiplier tube (PMT) readout system.

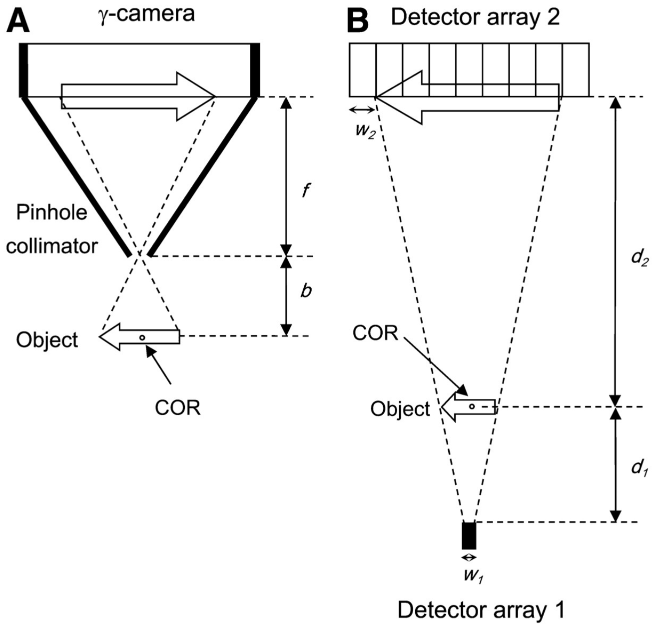


Figure 1.4: (a) Pinhole SPECT geometry; (b) Virtual pinhole PET geometry. COR denotes the center of rotation. (Reprinted by permission of the Society of Nuclear Medicine from: Y.-C. Tai, H. Wu, D. Pal, and J. A. O'Sullivan. Virtual-Pinhole PET. J Nucl Med. 2008; 49(3): 471-479, Figure 1.)

To improve sensitivity and to reduce costs, clinical whole-body PET scanners are usually built with medium-resolution detectors. The larger crystal size provides a larger interaction volume within each crystal (with a smaller proportion of dead space in between crystals in a block), and fewer crystals means fewer electronics channels are needed. High-resolution small animal PET scanners are often designed with a very small bore diameter to reduce the effect of acolinearity and to increase sensitivity (via larger solid angle) of the small crystals. As can be seen, there is a tradeoff between resolution and sensitivity.

A novel class of PET imaging systems was proposed by Tai et al. [96], which seeks to combine the advantages of small-animal imaging systems with the advantages of clinical systems. It is based on the theory of pinhole SPECT cameras, which use a lead pinhole collimator to acquire projections (as shown in Figure 1.4(a)). In the figure, b is the distance between the object and pinhole, and f is the distance between the pinhole and detector array. The projection of the gamma-emitting object onto the detector array is magnified by a factor of f/b. It is known that the resolution of a SPECT system depends more on the pinhole aperture diameter than the intrinsic resolution of the crystals, as long as the object is very close to the pinhole.

Figure 1.4(b) illustrates an analogous setup in PET. In this figure, detector 1 is higher-resolution than the crystals in detector 2. The object is projected onto detector array 2, this time using electronic collimation instead of physical collimation. In this setup, detector 1 can be thought of as a pinhole, hence giving the name of this concept virtual pinhole PET (VP-PET). There are two major benefits to such a system. The first is that degradation due to gamma ray acolinearity can be reduced, since there is a smaller distance between the detectors than there would be if the object were centered in a circular PET scanner. The second benefit is that the system resolution depends more on the intrinsic resolution of detector 1 than that of detector 2.

In Tai et al. [96], the VP-PET system resolution near the center of rotation was derived to be

$$R_{\mathrm{sys}} \approx \sqrt{\left(0.0088 \cdot \frac{d_1 d_2}{d_1 + d_2}\right)^2 + \left[\frac{d_2 w_1 + d_1 w_2 + |d_2 w_1 - d_1 w_2|}{2(d_1 + d_2)}\right]^2 + R_{\mathrm{src}}^2 + R_{\mathrm{block}}^2}, \tag{1.1}$$

where R_{src} is the resolution loss due to positron range, R_{block} is the resolution loss due to block effects, and d_1, d_2, w_1, and w_2 are as shown in Figure 1.4(b). As d_1 becomes small relative to d_2, the acolinearity (first) term can become smaller than the standard expression for R_{aco}, and the coincidence detection (second) term approaches w_1, the intrinsic resolution of detector 1.

When incorporated as an accessory “insert” device into an existing scanner, it can provide high-resolution imaging in the region near the insert crystals. It has the potential to offer image quality benefits beyond those of a scanner made entirely from high-resolution crystals, and at a fraction of the cost. VP-PET could be especially important in translational research (i.e., from pre-clinical to clinical), as there is a need for high-resolution imaging in humans to validate drug efficacy, etc. Finally, the VP-PET concept can be applied in any number of configurations, ultimately allowing for different types of applications.

The main disadvantage of a VP-PET system is that it significantly complicates the reconstruction process. This was cited by Chatziioannou [13] as being one of the main challenges with such a system, and we will see in Chapters 5 and 6 that this is indeed true.

1.4 Organization of the Dissertation

The general concept of image reconstruction from tomographic datasets is the same for both CT and PET, but the differing source/detector arrangement and physics necessitates algorithms specific to each. The focus of this dissertation work was the application of a general statistical image reconstruction framework to particular types of CT and PET scanners. This work has resulted in the development of several new reconstruction algorithms and efficient implementations that offer the potential for improved imaging performance over existing methods.

Chapter 1 (this chapter) was intended to provide a very brief overview of image science with an emphasis on the imaging modalities pursued in this work. Chapter 2 provides background material on various image reconstruction approaches for CT, PET, and virtual pinhole PET data. It also highlights the main contributions of this work. In Chapter 3, we describe a statical reconstruction method that has been applied to clinical multislice helical CT scanners. In Chapter 4, a modification of the CT reconstruction algorithm is presented that aims to reduce artifacts associated with incomplete datasets. Chapter 5 addresses the work that was done to reconstruct 2D PET insert data obtained from Monte Carlo simulations, and Chapter 6 addresses the work that was done to reconstruct 3D PET insert data measured by our physical prototype system. Chapter 7 provides the conclusions and some possible directions for future work.

Since reconstruction of CT and PET data is a complex task (especially for a novel system), there are many processing steps involved. It is recommended that the Table of Contents be used alongside the rest of this document to keep better track of how each section fits into the overall reconstruction framework.

Chapter 2

Background

In the context of medical imaging, a digital image is implicitly an approximate representation of some underlying continuous object (e.g., a patient). The rectangular region represented by each pixel (in 2D images) or voxel (in 3D images) is simply a convenient basis function, where the value in each pixel or voxel represents the strength of that basis function. In other words, a discretized image is a *model* for some physical object in continuous space, in which the value of each image element is a parameter of the model. A simple rectangular basis function does not allow for variation across each image element, and therefore the basis functions should be adequately small to capture any high-frequency variations. There are many alternatives to rectangular image elements, such as the "blob" model defined by Bessel functions [57], but we will not consider them to any great extent in this work.

Having explained the role of a digital image, we can now describe tomographic image reconstruction as the process of estimating a discretized image from a set of tomographic measurements. Reconstruction is a classic example of an inverse problem, in that it attempts to infer the model parameters describing an object from measurements of the object. The corresponding forward problem is that of predicting the data that would be measured given an object and the acquisition specifications. We will demonstrate in this chapter that modeling the forward problem is a key component in the design of reconstruction algorithms. This chapter will discuss some of the existing reconstruction algorithms used in CT and PET, and will also discuss two realizations of the virtual pinhole PET geometry.

2.1 Image Reconstruction Overview

Reconstruction algorithms for CT and PET can roughly be divided into two main categories: analytical and statistical methods. Analytical algorithms are based on a deterministic line integral model for the measured data, whereas statistical algorithms are based on an arbitrarily accurate model that also reflects the probability distribution of the measurements.

Both models apply to CT and PET data; however, in CT, the line integral approach is more commonly used, and in PET, the statistical model is more commonly used. This is because the count rate of photons at each measurement is usually several orders of magnitude greater in CT than it is in PET. The CT acquisition is therefore not count-limited in normal scan conditions, in which case a deterministic model works well. Limitations to patient dose and scan time, and the relatively low sensitivity in PET scanners (to detect all emitted gamma ray pairs) make the counting statistics in PET much poorer. Modeling the actual statistics of the PET data often improves the reconstructed image quality.

2.1.1 Reconstruction from Line Integral Data Model

The forward data model used by analytical methods is a line integral through the object along the path of a collimated x-ray beam (in CT) or LOR defined by two back-to-back gamma rays (in PET). As implied by the word *analytical*, the algorithm is derived in the continuous space; it is subsequently discretized for practical implementations. The following derivation is based on that of Kak and Slaney [42].

The derivation of the algorithm is simplest in parallel-beam geometry, shown in Figure 2.1. In the figure, t is the distance from the central ray, θ is the view angle with respect to the x-axis, and $\hat{\mathbf{n}}(\theta) \triangleq (\cos\theta, \sin\theta)$ is the unit vector specifying the perpendicular direction to the beam.[1] The Radon transform is defined as the parallel-beam

---

[1] Boldface notation will be used to represent vector quantities throughout the dissertation.

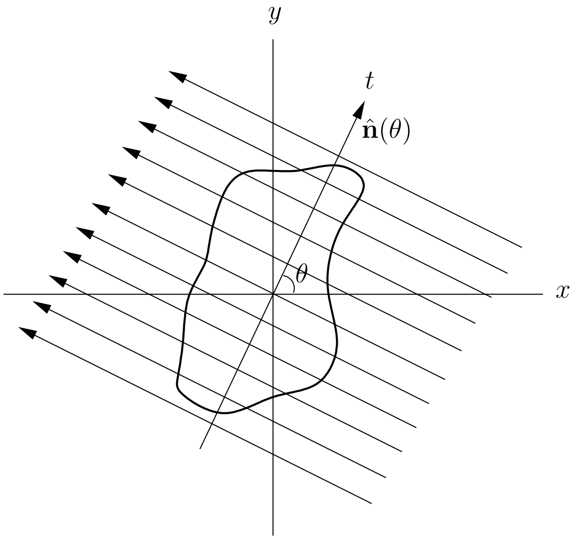


Figure 2.1: Parallel beam geometry used to define the Radon transform. The view angle is defined by θ, and a projection ray within a view is represented by t. The unit normal vector to the projection is given by $\hat{\mathbf{n}}(\theta)$.

projection through the object $f(\mathbf{r})$, where $\mathbf{r} = (x, y)$, and is mathematically given by

$$P_\theta(t) \triangleq \int_{\hat{n}\cdot r = t} f(\mathbf{r})\, d\mathbf{r} = \int \delta(t - \hat{\mathbf{n}} \cdot \mathbf{r}) f(\mathbf{r})\, d\mathbf{r}. \tag{2.1}$$

Taking the 1D Fourier transform of $P_\theta(t)$ with respect to t yields the Projection-Slice Theorem, which is the key to analytical reconstruction:

$$\begin{aligned} \hat{P}_\theta(\omega) &\triangleq \int_{-\infty}^{\infty} P_\theta(t) \exp(-j2\pi\omega t)\, dt && (2.2) \\ &= \int_{-\infty}^{\infty} \int \delta(t - \hat{\mathbf{n}} \cdot \mathbf{r}) f(\mathbf{r})\, d\mathbf{r}\ \exp(-j2\pi\omega t)\, dt && (2.3) \\ &= \int f(\mathbf{r}) \exp[-j2\pi\omega(\hat{\mathbf{n}} \cdot \mathbf{r})]\, d\mathbf{r} && (2.4) \\ &= \hat{f}(\omega\cos\theta, \omega\sin\theta). && (2.5) \end{aligned}$$

This shows that the Fourier transform of the projection at angle θ equals the 2D Fourier transform of $f(\mathbf{r})$ along the $\hat{\mathbf{n}}(\theta)$ direction.

We will now derive the fundamental analytical reconstruction algorithm for parallel-beam projection data, making use of the Projection-Slice Theorem along the way.

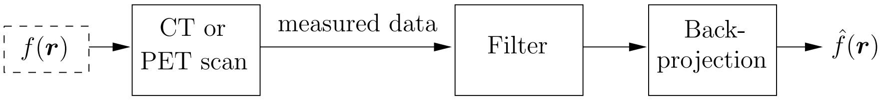


Figure 2.2: Linear FBP algorithm for CT and PET.

We start with the 2D inverse Fourier transform of the 2D Fourier transform of $f(\mathbf{r})$,

$$f(\mathbf{r}) = \frac{1}{(2\pi)^2} \int \hat{f}(\mathbf{k}) \exp[j2\pi(\mathbf{k} \cdot \mathbf{r})] \, d\mathbf{k}. \tag{2.6}$$

Now, we replace the rectangular frequency coordinates given by $\mathbf{k}$ with polar frequency coordinates, $\mathbf{k} \triangleq (\omega \cos\theta, \omega \sin\theta)$. The change of variables yields $d\mathbf{k} = \omega \, d\omega \, d\theta$, so (2.6) becomes

$$f(\mathbf{r}) = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{\infty} \hat{f}(\omega \cos\theta, \omega \sin\theta) \exp[j2\pi\omega(\hat{\mathbf{n}} \cdot \mathbf{r})] \omega \, d\omega \, d\theta. \tag{2.7}$$

Using symmetry arguments and the Projection-Slice Theorem, this can be simplified to

$$\begin{aligned} f(\mathbf{r}) &= \frac{1}{(2\pi)^2} \int_0^{\pi} \int_{-\infty}^{\infty} \hat{f}(\omega \cos\theta, \omega \sin\theta) |\omega| \exp[j2\pi\omega(\hat{\mathbf{n}} \cdot \mathbf{r})] \, d\omega \, d\theta & (2.8) \\ &= \frac{1}{(2\pi)^2} \int_0^{\pi} \int_{-\infty}^{\infty} \hat{P}_\theta(\omega) |\omega| \exp[j2\pi\omega(\hat{\mathbf{n}} \cdot \mathbf{r})] \, d\omega \, d\theta. & (2.9) \end{aligned}$$

Finally, we write this as

$$f(\mathbf{r}) = \frac{1}{2\pi} \int_0^{\pi} Q_\theta(\hat{\mathbf{n}} \cdot \mathbf{r}) \, d\theta, \tag{2.10}$$

where

$$Q_\theta(t) \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{P}_\theta(\omega) |\omega| \exp(j2\pi\omega t) \, d\omega. \tag{2.11}$$

From these two expressions, we observe that reconstruction of $f(\mathbf{r})$ entails high-pass filtering each parallel-beam projection using (2.11), and then backprojecting the result across the image plane using (2.10). The algorithm is known as filtered backprojection (FBP) for this reason. Figure 2.2 shows a flowchart of the FBP reconstruction procedure.

The high-pass filter is given by $|\omega|$ in the frequency domain, but should be apodized in practice to avoid amplifying high-frequency noise in the projection. It can be designed to control the noise-resolution tradeoff for different imaging protocols. The filtering operation can be performed very quickly in the frequency domain through the use of the fast Fourier transform (FFT).

An exact Radon inversion formula can be derived in a similar manner for the fan-beam geometry [42]. The resulting fan-beam algorithm is a weighted FBP formula. Another common alternative to using the fan-beam reconstruction algorithm is to use an approach that reorganizes, or rebins, the fan-beam projections into equivalent parallel-beam projections. Then, once the set of projections are in parallel-beam form, the standard FBP algorithm given by (2.10) can be applied. However, due to sampling pattern differences between the Cartesian and polar coordinate systems, interpolation is necessary in rebinning methods. Interpolation can adversely affect the noise-resolution tradeoff if not performed carefully.

In a similar manner, 3D data (i.e., data collected from multi-row CT scanners or multi-ring PET scanners) can be rebinned into approximately equivalent 2D fan-beam projections in transverse planes. However, the accuracy of the approximation depends on the obliqueness of the projection with respect to the axial direction. Some newer analytical reconstruction algorithms attempt to interpolate and reconstruct 3D data in more sensible oblique planes, as will be explained in more detail later.

2.1.2 Reconstruction from Statistical Data Model

Statistical reconstruction methods are generally based on maximum likelihood (ML) estimation or maximum a priori (MAP) estimation techniques. These require knowledge of the probability distribution from which the measured data arises. The Poisson distribution is a reasonable statistical model for CT and PET, although a compound Poisson model might be more appropriate for CT scanners with energy-integrating detectors [53].

In the case of Poisson random variables, the probability distribution for a particular measurement, d_i, is given by

$$P(d_i) = \exp(-g_i) g_i^{d_i} / d_i! \,, \tag{2.12}$$

where g_i is the mean of the measurement. Determination of the mean value of the measurement requires a model for the forward problem, known as the forward projection operation. Statistical algorithms for tomographic reconstruction rely heavily on forward projections, which at their simplest, are line integrals like (2.1). However, statistical algorithms are not constrained by the mathematics of the Projection-Slice Theorem or Fourier theory. A forward projection can instead be modeled by a general system matrix that relates the image space to the data space through a matrix-vector multiplication. It is given by

$$l_i(\boldsymbol{f}) \triangleq \sum_{j=1}^{N} a_{ij} f_j, \tag{2.13}$$

where N is the number of voxels in the image, a_{ij} is the system matrix, and f_j is the jth voxel of the image $\boldsymbol{f}$. Since the forward problem is modeled discretely from the outset, there is no need to perform interpolations if the system matrix is modeled appropriately. A description of the specific system matrix calculations will be deferred to Chapters 3-6.

In transmission tomography, the mean number of detected photons is governed by Beer's Law:

$$g_i(\boldsymbol{f}) \triangleq E[d_i] = I_i e^{-l_i(\boldsymbol{f})}, \tag{2.14}$$

where I_i is the mean number of photons generated and subsequently detected at source-detector pair i in the absence of an attenuating medium. A background term may also be added to (2.14), although we did not consider that in our current CT model; this is more important for transmission scans using a rotating positron source in PET scanners, due to the coincidence detection of unwanted randoms and scatter.

In emission tomography, the mean number of detected gamma ray pairs is equal to the forward projection plus some mean randoms and scattered counts:

$$g_i(\boldsymbol{f}) \triangleq E[d_i] = l_i(\boldsymbol{f}) + \bar{r}_i + \bar{s}_i. \tag{2.15}$$

ML estimation is an optimization problem that seeks to estimate the image parameters most likely to have given rise to the measured data [101, p. 65]. The likelihood function is given by $P(\boldsymbol{d}|\boldsymbol{f})$, and should be thought of as a function of $\boldsymbol{f}$ for the purpose of ML estimation. Mathematically,

$$\begin{aligned}
\widehat{\boldsymbol{f}} &\triangleq \arg\max_{\boldsymbol{f} \geq 0} P(\boldsymbol{d}|\boldsymbol{f}) && (2.16)\\
&= \arg\max_{\boldsymbol{f} \geq 0} \prod_i \exp(-g_i) g_i^{d_i} / d_i!, && (2.17)
\end{aligned}$$

where $\widehat{\boldsymbol{f}}$ is the ML estimate of the image, and the product is taken over all measurements. By writing (2.17) as a product of Poisson probabilities, it is assumed that each measurement is independent. It is easier then (yet completely equivalent[2]) to maximize the log-likelihood function, which is given by

$$L(\boldsymbol{d}|\boldsymbol{f}) \triangleq \ln P(\boldsymbol{d}|\boldsymbol{f}) = \sum_i d_i \ln g_i - g_i - \ln(d_i!). \tag{2.18}$$

The last term does not depend on $\boldsymbol{f}$, and therefore plays no role in the maximization process. From the definitions of $g_i(\boldsymbol{f})$ given in (2.14) and (2.15), it can be shown that $L(\boldsymbol{d}|\boldsymbol{f})$ is concave over $\boldsymbol{f}$ [24]. This guarantees that a (possibly non-unique) global maximum exists.

The ML solution is asymptotically unbiased (i.e., as the number of measurements approaches ∞), and provides the lowest-variance estimate of all unbiased estimators [101, p. 68–71]. Since tomographic systems have a finite limit to the number of measurements available, this fact does not necessarily carry over to real scanners. However, ML estimation offers a robust method in practice when the data is well-modeled.

In the Bayesian framework, MAP estimation can be used when the image parameters themselves are also random variables, i.e., they are generated from some underlying probability distribution. This "prior" distribution, $P(\boldsymbol{f})$, must be known in advance. Under this framework, the estimation procedure becomes that of maximizing the posterior distribution, which by Bayes rules is

$$P(\boldsymbol{f}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{f})P(\boldsymbol{f})}{P(\boldsymbol{d})}. \tag{2.19}$$

---

[2] Since the logarithm is a monotonically increasing function, $\ln P(\boldsymbol{d}|\boldsymbol{f})$ will have the same maximizer(s) as $P(\boldsymbol{d}|\boldsymbol{f})$.

By taking the logarithm of this expression, the MAP estimation problem becomes

$$\widehat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f} \geq 0}{\arg\max}\ L(\boldsymbol{d}|\boldsymbol{f}) + \ln P(\boldsymbol{f}). \tag{2.20}$$

If we let $P(\boldsymbol{f})$ be a Markov random field model, then

$$\widehat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f} \geq 0}{\arg\max}\ L(\boldsymbol{d}|\boldsymbol{f}) - \beta U(\boldsymbol{f}), \tag{2.21}$$

where β is a scalar weight determining the strength of $U(\boldsymbol{f})$ relative to $L(\boldsymbol{d}|\boldsymbol{f})$, and $U(\boldsymbol{f})$ is the Gibbs potential energy function,

$$U(\boldsymbol{f}) \triangleq \sum_{j=1}^{N} \sum_{j' \in \mathcal{N}_j} \phi_{j,j'}(f_j - f_{j'}). \tag{2.22}$$

Here, $\mathcal{N}_j$ is a local neighborhood of voxels surrounding voxel j, and $\phi_{j,j'}(\cdot)$ is often chosen to be a convex function. In this sense, the prior imposes some enforcement that the image be locally smooth, since $U(\boldsymbol{f})$ effectively penalizes large differences between neighboring voxel values. The prior expectation that images exhibit local smoothness is taken to be a reasonable assumption in CT and PET imaging. By varying β and choosing the function $\phi_{j,j'}(\cdot)$ appropriately based on the application, the resolution-noise tradeoff can be controlled over a wide range.

We take the point of view in this work that the $U(\boldsymbol{f})$ term in (2.21) is simply a penalty that leads to the construction of a penalized likelihood (PL) objective function; the image variables are not viewed as random variables originating from an explicit probability distribution. PL estimation can provide a degree of regularization when the reconstruction problem is particularly ill-posed. This avoids overfitting the noisy data to the likelihood function alone. It also benefits from the fact that the resulting image properties are determined by the objective function alone, not the algorithm used to optimize the objective function [24].

Statistical reconstruction uses iterative methods to optimize the objective function, as closed-form solutions to the optimization problem are not possible. Certain classes of algorithms are known to resolve lower-frequency components in earlier iterations than the high-frequency components. Depending on the convergence properties of the particular algorithm employed, stopping the iterations before convergence can offer an alternative practical form of regularization. It is more difficult to predict image quality in this case, but the implementation is simpler compared to MAP or PL reconstruction.

A linearized local impulse response function was derived based on the PL objective function in (2.21) and under the assumption of spatially-invariant scanner resolution [26]. Extensions to that work have resulted in approximate methods to predict the resolution, contrast, and variance properties of reconstructed images (without needing to reconstruct them) [74]. Unfortunately, these are mainly appropriate for quadratic penalties, which tend to over-smooth the image. Regardless of the limitations, Stayman and Fessler [91] provide the ability to systematically choose the β parameter as a function of desired resolution, which otherwise is very difficult to do.

Up until this point, we have presented a statistical reconstruction framework that is almost entirely identical for CT and PET. Since the particular form of $L(\boldsymbol{d}|\boldsymbol{f})$ is unique for transmission and emission tomography (due to the differing $g_i(\boldsymbol{f})$ functions), specific optimization methods to perform ML or PL estimation are needed for each modality. A brief overview is given below.

CT Statistical Reconstruction Algorithms

Many algorithms have been developed to optimize the objective function in transmission tomography. The methods shown in this paragraph update all image variables simultaneously. Lange and Carson [52] proposed an expectation-maximization (EM) algorithm for CT, although a closed-form update step could not be obtained, and convergence was slow. In Mumcuoğlu et al. [61], a preconditioned conjugate gradient algorithm was derived for the MAP estimator (for both CT and PET). This algorithm was shown to have a fast convergence rate compared to other algorithms. Another method was proposed that monotonically increases the penalized likelihood objective function at every iteration, even when there are background counts [22]. This method, as well as Elbakri and Fessler [21], used the concept of optimization transfer and paraboloidal surrogate functions. Finally, the alternating minimization (AM) algorithm proposed by O'Sullivan and Benac [66] provides a closed-form update for a more advanced data model, with guaranteed monotonic convergence.

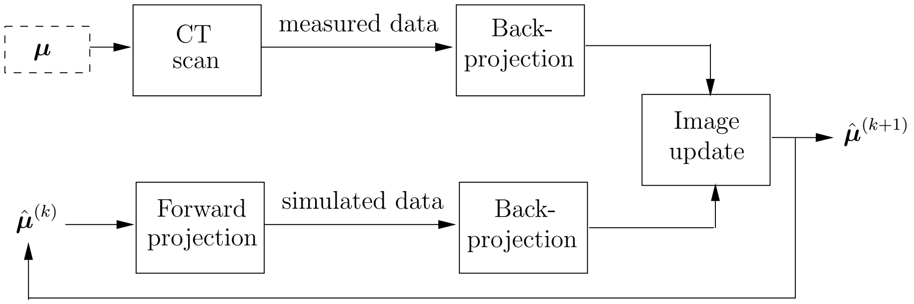


Figure 2.3: Iterative AM reconstruction algorithm for CT.

Another class of algorithms uses iterative coordinate descent, which entails updating each image variable sequentially [10]. It was shown to have a fast convergence rate, especially for high-frequency image components. In general, methods that update a larger number of variables simultaneously have slower convergence than methods that update one or few variables in parallel. This led to the grouped coordinate ascent method, which is a flexible approach that falls in between the fully parallel update methods and the fully sequential update methods [27].

Chapter 3 makes extensive use of the alternating minimization (AM) algorithm for transmission tomography [66]. A basic flowchart of this algorithm is shown in Figure 2.3.

PET Statistical Reconstruction Algorithms

Two independent derivations of the maximum-likelihood expectation-maximization (ML-EM) algorithm for emission tomography have been published [87, 52]. Both use the concept of a complete data space, which is a higher-dimensional space than that of the measured data alone, but makes the algorithm derivation easier. An alternative derivation is shown in Fessler [24] based solely on optimization transfer function principles. Figure 2.4 shows a flowchart of the ML-EM algorithm.

Since the ML-EM algorithm is a simultaneous update method, it has poor convergence properties. The preconditioned conjugate gradient algorithm [61] and iterative coordinate descent algorithm [10] mentioned above also apply to emission tomography, and improve the convergence rate compared to ML-EM. Another sequential

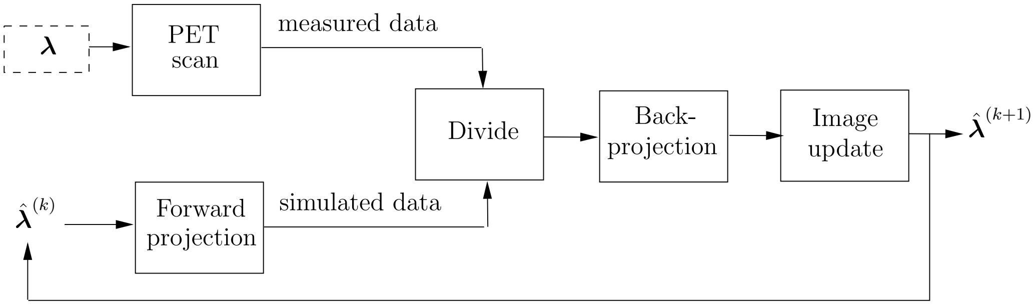


Figure 2.4: Iterative ML-EM reconstruction algorithm for PET.

update method with good convergence properties is the space-alternating generalized EM algorithm [25]. The interested reader is referred to the excellent review article by Qi and Leahy [75] for more information on statical reconstruction algorithms for emission tomography, which also includes a discussion of algorithms that process list-mode data directly.

2.1.3 Comparison of Analytical and Statistical Methods

Iterative algorithms are not usually implemented on commercial CT scanners because they are far too computationally intensive, and because the data collected often has enough counts that FBP algorithms can produce high-quality images. PET scanners generally use statistical reconstruction algorithms since the counting statistics are poorer, and would result in high-variance images if reconstructed with FBP.

Most CT scanners preprocess the measured data and output it in the form of line integral projection estimates, i.e., $l_i \approx -\ln(g_i/I_i)$. Taking the logarithm linearizes the measurements so they can be used with FBP-type algorithms. Fessler [24] showed that for random variables, the logarithm operation is systematically biased, especially for transmission measurements in the low-count regime. Statistical reconstruction methods do not use the measurements in projection space, so taking the logarithm of the transmission data is unnecessary.[3]

---

[3] However, since we only have access to the projection data on current CT scanners, we must first convert the measurement from the projection space back to the transmission space.

Additionally, FBP treats all measurements with equal weight since no noise model is used. This causes high-variance (i.e., low-count) measurements to be considered just as valid as low-variance measurements. Some methods attempt to reduce artifacts associated with high noise levels within the context of analytical reconstruction, such as Hsieh [33], but they are not as directly related to the measurement statistics as are methods that model the statistics upfront.

Despite the high computational cost, iterative methods for CT and PET are based on highly accurate models that take the underlying x-ray or gamma ray physics into consideration. For example, they can account for non-ideal scanning conditions such as noise, scatter, non-uniform detector response, x-ray beam hardening, and missing information due to low radiation dose or high-density attenuators (e.g., metal implants) in the field of view. Even in CT imaging, there are clinically relevant situations where the idealized analytical model breaks down (e.g., [105]). Statistical reconstruction algorithms were also cited by Wang and Yu [102] as being important for dose reduction strategies, and they predict this type of algorithm will be used more widely on commercial scanners once the computations become more manageable.

Fully 3D scanner geometries present a number of challenges for analytical algorithms. Scanner geometry is complicated by incomplete data sampling in some parts of the scan, redundant data in other parts, and oblique rays passing through the body. If not properly accounted for in the algorithm derivation for that particular scanner geometry, image artifacts can result. As mentioned above, statistical reconstruction algorithms use a general system matrix to represent the geometry and other effects; in this sense, the reconstruction algorithm remains the same regardless of the geometry.

Particularly within an iterative statistical reconstruction framework, there is considerable flexibility in terms of how the system matrix is designed to best model the measured data. Unlike analytical methods, which do not usually account for the finite spatial extent of each voxel and detector, the system matrix readily accounts for both. Modeling the voxels and detectors accurately provides the opportunity to better match the model to the intrinsic resolution of the system.

This dissertation is concerned with both the speed and accuracy of the reconstruction algorithms. Although there are significant computational benefits to using analytical reconstruction, we believe that the best accuracy can be achieved within the elegant statistical reconstruction framework. As will be shown in subsequent chapters, we have developed several strategies that can be employed to reduce the computation time of the statistical methods.

2.2 System Modeling

As seen in Section 2.1.2, statistical reconstruction algorithms have a similar structure in both CT and PET; at their core, they are built upon the use of forward projection and backprojection operations with a system matrix. The specific quantities that are represented in the system matrix are different for CT and PET, and will be detailed below.

2.2.1 CT Data

In the case of CT, the entries of the system matrix have the unit of length so that the argument of the exponential in (2.14) is unitless. Data from commercial whole-body CT scanners is usually precorrected for various effects, such as beam hardening and non-uniform detector response, so we will focus on using a simpler monoenergetic data model in this work. More sophisticated models exist, e.g., [50, 104].

There are many ways to forward project a volume of voxels into sinogram space, and conversely, to backproject data from sinogram space into the image space. De Man and Basu [17] provide a good review of some available methods in the literature, but for completeness, we describe a few of the methods here as well. The projection operations are similar to those used in ray tracing for computer graphics, except in that case, the rays are optical and therefore do not transmit through the object (unless transparent).

For forward projections, the most common method is the ray-driven approach; this is essentially the numerical calculation of a line integral through a voxelized space. A ray is drawn from the center of the x-ray tube's focal spot to the center of a detector element. The ray is then traversed at a given step increment, where at each stepping point along the ray, the voxels in its neighborhood are interpolated to that point. The interpolated result, multiplied by the step length, is then added to the projection value for the ray.

For backprojections, the most common method is the voxel-driven approach; this is how the backprojections in FBP-type algorithms are typically performed. For each voxel, a ray is drawn from the x-ray tube's focal spot to the center of the voxel. The ray is then extended further towards the detector array. If it intersects with the detector array, the nearby measurements on the detector are interpolated to the intersection point, and then added to the backprojection value for the voxel.

The calculations performed by these operations can be performed on-the-fly, but they can also be represented in a sparse system matrix. Even with more advanced interpolation methods, however, they do not accurately account for the finite size of the voxels and detector elements. They are also not transpose operations, as required by statistical reconstruction algorithms.

De Man and Basu [17] proposed distance-driven projectors as a more accurate method to perform the forward and backprojections. In 2D, the concept is to "splat" or project the voxel footprint and detector footprint onto a common axis, and then use the length of overlap as the weighting for that particular voxel and source-detector pair. In 3D, the footprints are projected onto a common plane, and the area of overlap is then used for the weighting. This approach does account for the finite extent of voxels and detectors, although some approximations are used to make the computations fast. In particular, the footprints that are projected to a common plane are approximated as being rectangular, even though they are actually a more complicated polygon that depends on the orientation of the voxel and detector relative to the ray.

In Politte and Whiting [72], a highly accurate projector for 2D fan-beam projections was proposed, which was also extended to 3D [73]. More details about this method are given in Section 3.2. Briefly however, it analytically or numerically integrates the length of intersection of each sub-ray with each voxel over the solid angle from the focal spot to the detector.

Lastly, a method was analytically derived to compute the system matrix for fully 3D circular-orbit cone-beam CT [109]. The method requires integration over different geometrical regions that correspond to portions of a voxel seen from a given projection angle. It has the advantage that numerical integration is not needed, so one does not need to calculate lengths of intersection over many sub-rays for sufficient accuracy. However, it does not apply to other scanner geometries, such as helical CT, and the implementation does not appear to be as straightforward as that of Politte et al. [73].

2.2.2 PET Data

The raw data from the PET scanner is usually saved in a list-mode file, which most importantly contains the detector pair index of each coincidence event. The list-mode data is not usually pre-corrected for any effects, as was the CT data. The only processing that is done prior to storage in the list-mode file is the hardware processing (e.g., coincidence timing window, energy discrimination). Although this puts a large burden on the reconstruction algorithm to model the data appropriately, it also provides the opportunity to model the main effects within a single statistical reconstruction framework.

To distinguish between the system matrix for CT and PET, we will now denote the PET matrix by P. In PET, each system matrix element p_{ij} represents a quantity that is proportional to the probability that a positron emitted in voxel j is detected by detector pair i. As done in Mumcuoğlu et al. [61] and as used in many other papers since then, a factored system matrix offers the ability to separately model various effects and to enhance the sparsity of the system matrix.

With respect to the basic model for PET data given by (2.15), where $l_i(\boldsymbol{f}) \triangleq \sum_{j=1}^{N} p_{ij} f_j$, P can be factored as [61]

$$P \triangleq P_{\text{geom}} \cdot P_{\text{norm}} \cdot P_{\text{atten}} \cdot P_{\text{blur}}. \tag{2.23}$$

P_{geom} is the geometrical factor that links the image space to the data space (in much the same way as the CT system matrix did). P_{norm} and P_{atten} are the diagonal normalization and attenuation factors, respectively, whose size in both dimensions is equal to the number of LORs. P_{blur} is a blur kernel used to model the positron range, gamma ray acolinearity, intercrystal scatter, and intercrystal penetration (if not modeled in P_{geom}).

The most important factor is P_{geom}, and it is most accurately calculated based on solid angle considerations. The relevant solid angle to use for this factor is the intersection of the following two quantities: the solid angle from the voxel to detector 1, and the solid angle from the voxel to detector 2 (which is subsequently extended backwards towards detector 1) [55]. This is the maximum solid angle in which an emission from a given voxel can be detected in coincidence by the detector pair. In practice, the voxel should be subdivided into smaller point-like sources to find the effective solid angle for the whole voxel and detector pair [76, 55]. To simplify the calculations, these two references assume that the inherently 3D solid angle can be approximated by the product of the 2D angle in the transverse direction and the 2D angle in the axial direction. This holds when the area of the crystal front face is very small compared to the squared distance between the emission point and the solid angle-limiting detector.

For whole-body PET scanners, the solid angle can normally be calculated using the front faces of a detector pair if intercrystal penetration will be modeled separately in P_{blur}, or if intercrystal penetration is not expected to lead to significant resolution degradation. For small-animal or dedicated organ-specific imaging systems, there has been recent interest in developing analytical detector response functions (DRFs) that explicitly calculate the intercrystal penetration effect within the geometrical computations, e.g., [94, 108]. These were designed to model depth-of-interaction (DOI) detectors. The main concept is shown in Figure 2.5, from Yamaya et al. [108], which modeled the transverse direction. Using their notation, they write the DRF for a point source located at $\mathbf{r}$ and detected by the ith DOI layer index of detector pair j, as

$$\begin{aligned} h_{i,j}(\mathbf{r}) = \frac{1}{2\pi}\int_0^{2\pi} & \{1-\exp[-\mu l_{(i,j),A}(\Omega,\mathbf{r})]\}\{1-\exp[-\mu l_{(i,j),B}(\Omega,\mathbf{r})]\} \\ & \cdot \exp[-\mu \bar{l}_{(i,j),A}(\Omega,\mathbf{r})]\exp[-\mu \bar{l}_{(i,j),B}(\Omega,\mathbf{r})]\, d\Omega, \end{aligned} \tag{2.24}$$

where μ is the attenuation coefficient of the detector crystal at 511 keV. This expression models the crystal interference lengths, $\bar{l}_{(i,j)}(\Omega,\mathbf{r})$, as well as the probability of detection inside the layer of interest, based on the lengths $l_{(i,j)}(\Omega,\mathbf{r})$.

As shown in Strul et al. [94], analytical DRFs are not as useful for large-diameter systems since other factors like gamma ray acolinearity play a larger role. As the methods are also very computationally intensive – especially in 3D – they have not

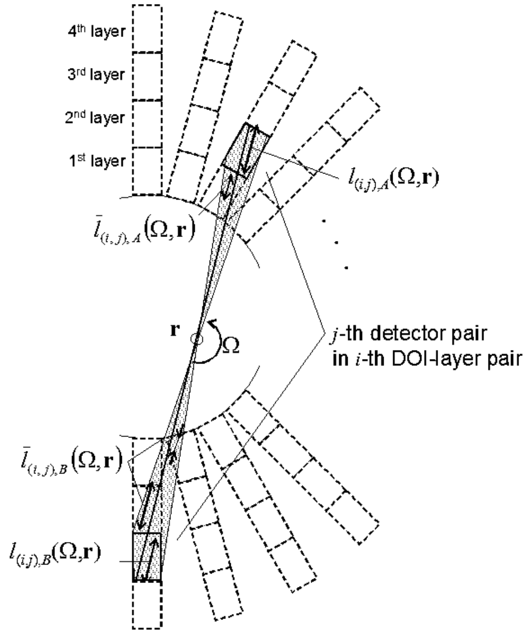


Figure 2.5: Geometry and notation for the analytical DRF model. (Reprinted by permission of IOP Publishing Ltd from: T. Yamaya et al. Transaxial System Models for jPET-D4 Image Reconstruction. Phys Med Biol. 2005; 50: 5339-5355, Figure 1.)

been used to model whole-body scanners. The computations are most efficiently performed by iterating over each voxel in the outer loop, and then iterating over all possible detector pairs that see the voxel in the inner loop. For sequential-update reconstruction methods, storing the system matrix in a lookup table in this order makes sense. However, this ordering would not work well during parallel-update reconstruction methods (such as ML-EM), since the reconstruction algorithm would not be able to use the cache efficiently.

Less computationally demanding (but also somewhat less accurate) methods have also been developed to model the solid angle with intercrystal penetration effects, e.g., [37, 59]. These methods are based on subdividing the crystals into subvolumes, and calculating the length of intersection of the sub-rays and voxels. Moehrs et al. [59] used Gauss-Legendre quadrature for the optimal integration points, whereas Huesman et al. [37] used a regular grid spacing within each crystal volume for the integration points.

The attenuation and normalization factors will be discussed further in Chapters 5 and 6. In Chapter 6, we will discuss scatter and randoms estimation in more depth as well. We will not consider the P_{blur} factor in this work. It can be calculated using small Monte Carlo (MC) simulations that assume rotational symmetry and spatial invariance [76]. The blur kernel is essentially a sparse banded diagonal matrix, where the size of the band is related to the extent of local blurring to other crystals.

It should be mentioned that other techniques can also be used to determine the PET system matrix, with varying degrees of effort. Some groups have employed MC simulations to calculate the entire system matrix without analytical calculations [77, 78]. The system can be modeled quite accurately in this manner, at the expense of computation time to simulate enough counts. Assuming there are no alignment issues, the least biased method is to collect point source measurements on the scanner itself. Using a robot to make the measurements, along with symmetry and model-fitting techniques, this was done in a practical amount of time for a clinical whole-body PET scanner [71]. They demonstrated superior image quality over conventional system matrix generation methods.

2.3 Helical CT Reconstruction

Multislice helical x-ray CT has proven to be a successful imaging modality in many clinical applications, and is now in widespread use. This type of CT is inherently 3D because the x-ray tube continuously projects a cone-beam through the patient as the patient is translated through the scanner. Each detector row captures data in a partial rotation of the gantry that corresponds to each image slice. Practically all helical CT scanners use some form of analytical image reconstruction algorithms based on filtered backprojection (FBP). Such algorithms are fast and usually yield high-quality images.

Significant progress has been made in the development of analytical reconstruction algorithms specific to helical CT. A number of techniques based on PI-lines [15] have been developed for exact reconstruction [44, 63, 114]. Several heuristic FBP-type algorithms (known as approximate algorithms) have also been developed that offer more practical implementations. Among these, the adaptive multiple plane reconstruction method rebins the helical CT data into oblique planes, upon which 2D

FBP is performed; the reconstructed tilted slices are then interpolated into an image volume with parallel slices [82].

Helical FDK is another well-known approximate method, in which weighted FBP is performed using 3D backprojection. Tang et al. [97] account for redundancy in the helical measurements through a 3D weighting function based on the concept of complimentary (also called conjugate) rays. These are rays that pass through the same voxel and are co-linear when projected onto the transverse plane. Although the method is somewhat empirical, it was shown to produce high-quality images under ordinary scan conditions. Stierstorfer et al. [93] propose another weighted FBP approach, where in this case the weights associated with conjugate rays can be calculated as part of the overall algorithm instead of requiring an explicit, possibly complex, weighting function. However, they also recommend using a row-dependent weighting function (which serves a different purpose).

The fast-paced advances in CT data acquisition hardware are making existing image reconstruction algorithms inappropriate or impossible to use, however. For example, since cone beam projection data is acquired in oblique planes, errors are often introduced as a result of interpolation to the transverse plane. As the number of detector rows increases, as is expected in future CT scanners, the interpolation errors will continue to grow; a wider detector array in the axial direction requires a cone beam with larger solid angle, thereby resulting in more oblique projections relative to the transverse plane.

Due to sampling limitations inherent to the cone-beam geometry, an overscan length is required on each side of the desired axial FOV. Some analytical reconstruction methods for helical CT cannot make full use of the overscan data, thereby resulting in an unnecessary increase in radiation dose to the patient. Some of the most state-of-the-art CT scanners (e.g., the Siemens Somatom Definition Flash scanner and the Toshiba Aquilion/Premium scanner) now have an active collimation system (pre-patient) to shape the x-ray beam to the detectors whose measurements will actually be used in the reconstruction.

Unlike some of the above methods, statistical iterative reconstruction algorithms inherently use 100% of the acquired data, which are weighted according to the statistics of the measurements. Additionally, statistical methods do not depend explicitly on the scanner geometry. Instead, the scanner is modeled by a general system matrix that describes the relationship between the discretized image and data spaces. The matrix elements are designed to account for factors such as finite detector aperture, finite voxel size, and focal spot blurring.

Very little work has been published on implementations of statistical reconstruction algorithms applied to helical CT scanners. The first such study was done within a maximum likelihood (ML) framework for single-slice helical CT using simulated parallel-beam projections [64]. Our group has previously described an iterative reconstruction implementation for multislice helical CT [73], also using ML principles. Due to its large computational requirements, significant data downsampling was needed to generate results using the non-optimized serial implementation.

Another group has worked on iterative reconstruction using the helical CT geometry, with an emphasis on region of interest reconstructions [113, 38]. Their main application was cardiac imaging, where the region of interest is very small compared to whole-body CT. For applications where the region of interest is small, their methods provide a practical means to perform high-resolution 3D iterative reconstruction without requiring a very large number of voxels. As such, their work did not focus on the computational aspects of 3D iterative reconstruction for whole-body CT imaging.

A method that uses iterative coordinate descent (ICD) on a penalized, weighted least squares objective function was also recently proposed [99]. Through detailed image quality analysis, they found significant improvements in resolution and artifact reduction using ICD over FBP-type methods. Since they used the quadratic approximation to the Poisson log-likelihood function – which leads to bias at low photon count rates – this method may be sub-optimal for low-dose scanning. Additionally, ICD is not readily parallelizable, unless for example, a grouped coordinate descent algorithm is used [27]. Their results were therefore necessarily limited to small scan volumes.

One further method with questionable application to statistical reconstruction of helical CT data was proposed [92]. The main point of the paper was to show that image rotations could lead to a significant amount of symmetry, namely, the elimination of axial dependence on the weights. However, this method was only tested with an analytical FDK method, and is unlikely to perform well within the context of statistical reconstruction due to the requisite use of a very simple line integral model.

Additionally, image rotation could lead to degraded resolution and difficulties in the appropriate application of regularization.

2.4 Incomplete CT Data Reconstruction

One important problem in CT and PET is that of missing or incomplete data. In this work, we will only focus on the CT problem, however.

When the support of an attenuating object lies partially outside the field of view (FOV) of a scanner, artifacts may arise in the reconstructed image because the object is undersampled. Most reconstruction algorithms implicitly assume the entire object is confined to the FOV, but if this is not the case, excessively large attenuation values may be reconstructed inside the boundary of the FOV. This observation results from the attempt to reconstruct an image entirely within the FOV such that projections through it will match the measured data; however, the reconstruction algorithm is unaware that the measured data has been affected by the object's attenuation outside the FOV.

Common to both 2D and 3D imaging is the issue of transverse truncation, which can arise when some part of the object – either deliberately or unintentionally – is not fully contained within the FOV in the transverse plane. In some cases, for example, it is desirable to scan just a region of interest (ROI) as a means of limiting the dose to a radiation-sensitive organ. Unintentional truncation may occur when imaging obese patients or when the bed does not fit in the FOV. In PET/CT scanners, patients often leave their arms at their side for comfort during the PET scan. The CT scanner FOV is not necessarily large enough to encompass their arms, which could then lead to CT artifacts that propagate into the attenuation correction for PET reconstruction. As a final example, during radiation treatment planning, patients are imaged in a CT scanner with bulky immobilization devices that often do not fit in the FOV of a general-purpose CT scanner.

Reconstruction methods have been specifically designed to reduce artifacts due to transverse truncation. As mentioned above, in applications such as attenuation correction for PET reconstruction, it is important to obtain quantitatively accurate CT images that cover the entire extent of the object in each slice. Methods that attempt to solve this problem reconstruct the image within the FOV as well as in the extended FOV – the region between the bore of the scanner and the actual FOV. Hsieh et al. [34] and Sourbelle et al. [90] extrapolate the missing data in each projection using 2D parallel beam consistency conditions (e.g., constant area under the projection curve for each view) and other constraints. In practice, methods that first extrapolate the missing data can use either analytical or statistical reconstruction algorithms on the modified sinogram; however, it is not clear what effect this might have on the data statistics if using statistical reconstruction.

Among other approaches are wavelet-based multi-resolution analysis [86] and total variation minimization image reconstruction [88]. To address the limited data problem in radiotherapy online CT systems, Ruchala et al. [79] proposed aligning and merging the data from a planning CT scan (used as prior information) with the truncated data collected by an online CT system at the time of therapy.

Modern multislice helical CT scanners cannot fully sample the beginning and end slices of a scan since the x-rays are collimated into a wide cone-beam originating from a single x-ray tube. In the axial direction, this implies that an incomplete set of projection data will be collected when scanning a portion of a long object. Unless an overscan is employed, it is currently difficult to image a patient without z-axis undersampling in the ROI. In recent years, the so-called long object problem has been of considerable interest to those developing analytical reconstruction techniques for multi-slice spiral CT and cone-beam CT, e.g., [49, 81, 19, 114]. It appears that less work has been reported on this problem using statistical reconstruction approaches, most likely due to the fact that statistical methods are not specific to a particular system geometry.

The analytical and statistical methods are typically quite different from one another, although there have been recent attempts to relate analytical data sufficiency conditions to iterative reconstruction methods in emission tomography [29, 112]. Additionally, Defrise et al. [20] have developed a sufficiency condition that depends on the inversion of the truncated Hilbert transform; since the inversion formula is not known analytically, they used MAP reconstruction to evaluate the sufficiency condition.

Within the context of analytical reconstruction, a number of papers have specifically addressed the long object problem, which can be considered a type of ROI tomography. For example, Schaller et al. [81] used an exact rebinning method called the PHI-method, while other groups have made use of differentiated backprojection along PI-line segments [19, 114].

Zeng et al. [111] published a method that uses iterative reconstruction to address the long object problem. It is similar in principle to the first approach of Snyder et al. [89] (with one major exception being the choice of reconstruction algorithm). In Zeng's method, rays that pass through both the ROI slices and outside slices of the object are not used. More recently, La Rivière [50] described a method that performs statistical reconstructions within the FOV from truncated data. An initial estimate of the projections outside the FOV is obtained by computing an analytical reconstruction inside the FOV, and then subtracting its reprojection from the measured projections. The projection estimates are iteratively refined along with the pixel values within the FOV through a joint estimation procedure.

Finally, the two methods presented in Snyder et al. [89] are general, straightforward extensions of the AM reconstruction algorithm described previously. They can be applied to any system geometry, and are therefore suitable candidates for investigating transverse and longitudinal truncation. In particular, we will demonstrate in Chapter 4 that the missing data problem can be addressed directly in 3D without needing to rebin projections to simpler system geometries. For example, this could potentially allow for extended FOV reconstructions without using consistency conditions specific to the parallel beam geometry.

2.5 Virtual Pinhole PET Systems and their Reconstruction

In addition to the proof-of-principle setup shown in Figure 1.4(b), our group has developed two prototype VP-PET systems that integrate with existing scanners. The first is a full-ring insert for a microPET F220 scanner (Concorde Microsystems, Knoxville, TN), and the second is a half-ring insert for a clinical Siemens Biograph 16 PET/CT scanner (Siemens Medical Solutions, Knoxville, TN).

2.5.1 Full-Ring Insert

The full-ring insert was developed to enhance the resolution of existing small-animal imaging systems. It consists of a ring of high-resolution scintillation crystals oriented inside the boundary of the scanner FOV. Given the small radius of the microPET scanner, it would be impossible to fit the PMTs and readout electronics behind the insert crystals (within the bore of the scanner). Instead, fiber optics are used to carry the scintillation light from the crystals to PMTs located outside the scanner FOV. A side benefit of this design is that it reduces the overall amount of attenuation and scatter compared to the (impossible) case where the insert PMTs and readout electronics are coupled directly behind the insert crystals in the FOV.

Since the entire imaging FOV is inside the insert ring, the geometry of this system resembles that of a fourth-generation x-ray CT scanner. Three types of coincidence are allowed: between two scanner crystals (SS), between two insert crystals (II), or between one insert crystal and one scanner crystal (IS). The II data is the noisiest, but has the highest resolution, and the opposite is true of the SS data; IS falls in the middle. A 2D FBP algorithm was derived to handle the asymmetric insert-scanner coincidences [70]. In the same work, a more robust reconstruction method was developed, based on a variation of the ML-EM algorithm, that utilized all three types of coincidence. The ML-EM approach was also extended to 3D [68].

2.5.2 Half-Ring Insert

To translate the virtual pinhole PET concept to clinical PET scanners, a second type of insert geometry was proposed. Rather than use a full-ring insert as in the small-animal imaging system above, a half-ring insert design was introduced to accommodate a larger range of object shapes and sizes. The half-ring insert gantry contains detector modules that are arranged such that they form the lower half of a circle, with the circle's center aligned with the center of the PET scanner's FOV.

Two main directions of study have been carried out for this geometry, which will be briefly reviewed below. Much of this work will be the focus of later chapters in this dissertation. The half-ring insert was first studied using Monte Carlo simulations, where the main application was breast cancer imaging. A physical half-ring prototype insert system has also been developed, and is primarily designed for high-resolution head and neck cancer imaging. High-resolution breast cancer imaging is another potential application for the physical prototype, although that is beyond the scope of this work.

With regard to breast cancer imaging, many groups have built dedicated PET scanners for breast cancer detection (known as positron emission mammography) that offer higher spatial resolution and higher sensitivity than a conventional whole-body PET scanner. These dedicated systems use detectors that surround (or rotate around) the breast. However, one of the main limitations of this approach is that the scanner geometry precludes the imaging of the chest wall and axillary lymph nodes since these regions are typically outside the scanner FOV [28]. Lesions are commonly found in these regions as well as in the breast tissue, and therefore it is important to be able to image them in addition to the breast. Our half-ring insert design overcomes this limitation by making it possible to image the breast with higher resolution than the standalone scanner, while still maintaining the FOV that includes the chest wall and the rest of the thorax. A similar half-ring system for breast cancer imaging was proposed in Tai et al. [95].

Our physical prototype system for head and neck cancer imaging was first described in Wu et al. [106]. The insert device is centered in the clinical whole-body PET scanner, and connected to the electronics within the scanner. The patient bed has a narrow extension at the end closest to the scanner for the patient's head to rest during the scan. This extension, unlike the bed itself, fits into the FOV of the insert device. If this prototype system were to be used for breast cancer imaging, the patient bed and other components of the insert system would need to be modified.

Irrespective of the application, the insert can be thought of as an accessory device for a whole-body PET scanner, providing a high-resolution FOV without the need for a dedicated PET scanner. The use of a half-ring insert inside a whole-body scanner enables the acquisition of II, IS, and SS coincidences like in the full-ring insert system. Since the II and IS datasets provide incompletely sampled data in the half-ring geometry, and all three coincidence types are of different resolution, there are a number of unique challenges posed by this system. For example, FBP reconstruction algorithms cannot be used due to the incomplete sampling and our desire to jointly use all three datasets to form a single image. We have therefore chosen to take advantage of the flexibility offered by the statistical reconstruction framework.

At a high level, the half-ring ML-EM reconstruction procedure is similar to the algorithm used for the full-ring insert system, although the half-ring insert requires more advanced system modeling. This is partly due to the fact that the side faces of the insert crystals are also exposed to gamma ray radiation, whereas in the full-ring insert system, this was not the case. Initial work on the 2D reconstruction of simulated half-ring insert data was described in Pal et al. [69]. We will improve upon that initial modeling effort in Chapter 5, and extend it to the real 3D system in Chapter 6.

2.6 Acceleration of Statistical Reconstruction Algorithms

Most of the time spent in statistical iterative reconstruction algorithms is in the computation of the forward projections and backprojections. Considering the many benefits of statistical reconstruction, one goal in the research community is to speed up these methods to clinically feasible computation times. A variety of acceleration techniques have been developed, and can be divided into algorithmic and hardware approaches.

2.6.1 Algorithmic Speedup

Ordered subsets (OS) were introduced in Hudson and Larkin [36], and are able to speed up the convergence of parallel-update iterative reconstruction algorithms significantly. An algorithm that utilizes OS iteratively computes image updates using only a subset of the available projection data. During each iteration, the OS algorithm cycles through each subset of data, performing an image update after each sub-iteration. OS is able to improve the convergence rate by a factor roughly equal to the number of subsets, since

$$\nabla_{\boldsymbol{f}} L(\boldsymbol{d}|\boldsymbol{f}) = \nabla_{\boldsymbol{f}} \sum_i (d_i \ln g_i - g_i) \approx M \cdot \nabla_{\boldsymbol{f}} \sum_{i \in \mathcal{S}_m} (d_i \ln g_i - g_i) \tag{2.25}$$

(at least in the early iterations). Here, M is the number of subsets, and $\mathcal{S}_m$ is the subset of measurements corresponding to subset index m. There is negligible overhead with ordered subsets; the computation time for each sub-iteration is simply scaled down by the number of subsets.

The original OS method removes the monotonic convergence guarantee of most statistical reconstruction algorithms. Convergent OS methods have been developed [1], but their memory demands may be too high for clinical practice. However, even without the convergence guarantee, the original OS method tends to be stable in practice.[4]

A different class of approaches that are specific to 3D PET entail first rebinning the data into 2D sinograms, and then performing iterative reconstruction (or FBP) in 2D. The simplest way to do this is to use the single slice rebinning algorithm (SSRB) [16], although this is only accurate when the activity is near the isocenter. A better method is to use Fourier rebinning (FORE) [18], in which the obliqueness of the rays is properly accounted for in the rebinning. This method is an approximate discretization of an exact inversion formula (based on line integrals). Another rebinning approach makes use of both FORE and inverse FORE [14]. In this method, the forward and backprojection operations are performed quickly in 2D, but in between each iteration, the data is converted back into fully 3D sinograms so 3D blur kernels and other system matrix factors can be applied. This method offers the potential for some additional resolution recovery over FORE alone.

Another method for CT and PET is the hierarchical decomposition strategy of Basu and Bresler [7], which reduces the computational complexity of the forward and backprojections in a manner similar to the decomposition of the discrete Fourier transform

---

[4] The chosen number of subsets cannot be too aggressive, and the subsets must be approximately balanced.

into the fast Fourier transform. Symmetry can also be exploited in the system matrix, depending on the scanner geometry and image space. We discuss symmetry in our CT and PET reconstruction algorithms in subsequent chapters. An interesting alternative to rectangular voxels is the use of polar voxels as the image basis functions [60]. For PET scanners, this could lead to a significant increase in transverse symmetry (under a few assumptions). Unfortunately, to display such voxels on a computer screen, scan conversion would be needed to reformat the image to rectangular voxels; this process could potentially lead to resolution degradation. The system matrix would also be difficult to compute analytically, and might therefore require lengthy MC simulations.

2.6.2 Hardware Speedup

Multicore and multiprocessor implementations of 3D statistical reconstruction algorithms are becoming increasingly necessary to keep up with the massive amount of data that needs to be processed from the scanner. Parallelized implementations have been reported in the literature for certain modalities. Examples can be found in PET [39, 40, 100, 84] and circular-orbit cone-beam CT (CBCT) [47, 9], for example. Analytical and statistical reconstruction algorithms for circular-orbit CBCT have also been implemented on the Cell Broadband Engine to perform extremely fast reconstructions [41, 46].

Implementations for PET and circular-orbit CBCT do not need to account for movement along the axial direction of the scanner during data acquisition, as they would need to do in helical CT. This added challenge, along with the fact that very few articles have been published on helical CT statistical reconstruction in the first place, could explain why there is no existing literature on parallelized statistical reconstruction methods for helical CT. We will revisit this topic in Chapter 3.

Another emerging trend is the use of graphical processor units (GPUs) in image reconstruction. These are by far the least costly option in terms of parallel computing power, and they can provide large speedups over single-CPU reconstruction implementations due to their specialized ability to handle graphics operations efficiently [107, 48]. However, it is currently difficult to fit the full data and image arrays into the limited graphics card memory. There are high latency penalties if the GPU has to utilize the external memory bus to retrieve data. Additionally, GPUs cannot execute branches (“if” statements) unless all processor cores take the branch, whereas CPUs are much better at this type of control flow. A modification to the distance-driven projector mentioned above was proposed that eliminates branching [8]. In general though, despite significant advances in the programming of GPUs (e.g., Nvidia CUDA), it is still not easy or even possible to compute highly accurate system matrices on GPUs.

As seen by the variety of implementations that exist for 3D image reconstruction, there are clearly significant tradeoffs that one must make between writing fast, highly optimized code for a particular architecture and writing slower, portable software that can be used in a more flexible manner. For certain applications, such as CBCT, GPUs are a natural choice since the geometry of flat-panel detector cone-beam systems is essentially the same as that for which GPUs were originally designed. However, implementing multi-row helical CT reconstruction on a GPU would be less straightforward because of the curved detector design and continuous patient feed through the scanner.

2.7 Main Contributions

The primary focus of this work was to apply general statistical image reconstruction concepts to specific clinical imaging modalities and scanner geometries. In the case of multislice helical CT, only a handful of papers appear in the literature describing fully 3D statistical reconstruction algorithms, and none of them focus on being able to process large clinical datasets; most research pertaining to this scanner geometry has focused on analytical approaches due to their relatively low computational cost.

In the case of virtual pinhole PET, our group is a pioneer in developing this novel imaging geometry for PET. Many of the reconstruction methods found in the literature for PET imaging do not apply directly to our novel insert system, and had to be redeveloped by the author and lab members. This work required extensive understanding and modeling of the data acquisition process. Through detailed comparison with simulated and measured data, the author developed and refined appropriate models and normalization methods that could be used to estimate the data and image during the course of the reconstruction process.

Given the enormous computational requirements of fully 3D statistical reconstruction algorithms compared to most commercially-implemented methods, it was necessary to develop custom parallelized approaches and other forms of optimization within the 3D reconstruction code. For example, a new encoding of the helical CT system matrix lookup table was developed to ensure that it would fit within the memory limitations of a workstation, and that it would lead to faster reconstructions. Specific forms of symmetry were identified and implemented in the helical CT and PET insert system code as well.

Some of the basic ideas that went into the helical CT code were based on previous work by Shenyu Yan and Dr. David Politte, although the code itself was written completely independently. Even without parallelization, the optimized code for generating the system matrix is orders of magnitude faster than a previously-written version. The reconstruction framework that was developed by the author has enabled our group to reconstruct clinical data using fully 3D statistical reconstruction algorithms for the first time. Another contribution to the helical CT work was the addition of a non-quadratic penalty term to the objective function used by the AM algorithm. The helical CT data was acquired with the help of Dr. Bruce Whiting.

For the MC simulations of the half-ring insert system, the input scripts were initially based on work by Dr. Martin Janecek and Dr. Sergey Komarov. Over time, the author wrote other advanced MC scripts and methods to process the raw output for use in the reconstruction code. The prototype half-ring insert was designed and built primarily by Dr. Heyu Wu, Dr. Yuan-Chuan Tai, and Dr. Tae Yong Song. Dr. Komarov implemented a scatter simulation technique whose input and output files are based on the author's reconstruction code specifications. The method that was implemented by Dr. Komarov is described in Section 6.2.4.

The author traveled with the group to Siemens (Knoxville, TN) on multiple occasions to be involved in the hardware integration of the insert with the whole-body PET scanner. Code was written by the author to communicate certain hardware settings, such as the peak positions in flood images and crystal energy photopeaks, to the Siemens scanner firmware. Additionally, custom list-mode sorting code was written by the author to process the raw encoded output from the scanner into fully 3D II, IS, and SS detector matrices for the prompt counts and delayed counts. Based on this code, the author helped develop a procedure to align the insert within the scanner based on multiple point source acquisitions. He has also been actively involved in the design and acquisition of all experiments, and originally proposed the idea of acquiring a separate attenuation map of the insert for normalization. He led the efforts in obtaining a useful transmission image of the insert. It is expected that the reconstruction system developed by the author for the half-ring insert will be used for all future reconstructions of data acquired with this prototype. The original methodology that was developed should also be useful in future virtual pinhole PET designs.

Chapter 3

Reconstruction of Multislice Helical CT Datasets

In this chapter, we present the details of the efficient fully 3D reconstruction framework that has been developed for helical CT. Section 3.1 describes the scanner geometry and statistical iterative reconstruction theory. Section 3.2 derives an accurate expression for the system matrix elements, and discusses the symmetry inherent to the helical CT geometry. In Section 3.3, we present our parallelization method. Section 3.4 explains the experiments we have conducted to validate our approach, while Section 3.5 provides these results. We discuss the results in Section 3.6.

3.1 Theory

Statistical reconstruction algorithms require an accurate model of the system geometry and the data that is to be collected. Below, we highlight the main aspects of these components, along with a formulation of the reconstruction algorithm.

3.1.1 System geometry

The geometry of the multislice helical CT scanner is shown in Figure 3.1. The focus of the x-ray source rotates at a radius of R_f from the isocenter, and the detector array, described by a circular arc of radius $(R_f + R_d)$, rotates in the same direction at a radius of R_d from the isocenter. Using continuous coordinates, β is the angle between

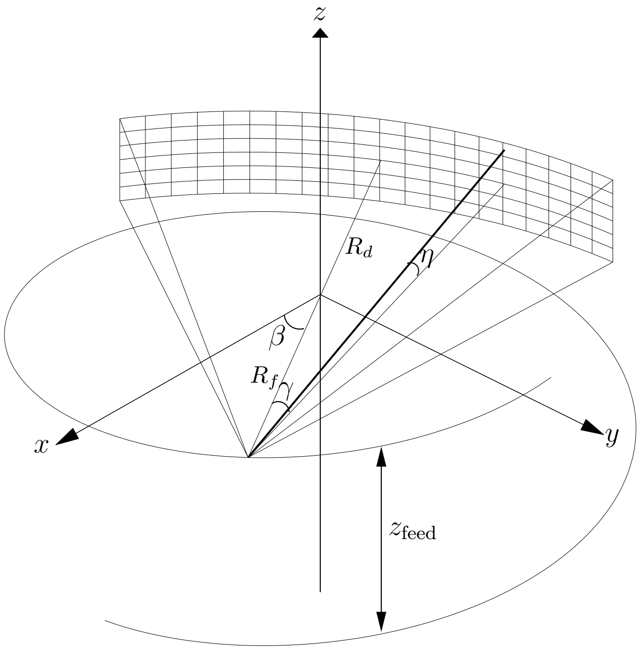


Figure 3.1: The multislice helical CT geometry used in this chapter.

the positive x axis and the line connecting the focal spot to the isocenter. This angle is not restricted to $[0, 2\pi)$, but rather, continues increasing with subsequent gantry rotations. A particular location on the curved detector array is specified by the fan angle, γ, and the cone angle, η. Finally, z_{feed} is the axial distance traveled by the patient bed (or equivalently, the gantry) in one rotation.

3.1.2 Statistical data model

Detailed data models that account for scatter, noise, and beam hardening exist in the literature [24, 104, 53, 66]. In this chapter, we consider a mono-energetic, scatter-free statistical model that accounts for the randomness in the x-ray photons. Such a model can capture the x-ray physics and scanner geometry reasonably well, while also simplifying the computations in the fully 3D reconstruction problem to be described.

The basic statistical model for the measured transmission data, d_i, assumes the photons arrive at the detector elements according to a Poisson counting process. The index i refers to a particular source-detector pair. In discretized form, the mean value of d_i is modeled as

$$g_i(\boldsymbol{\mu}) \triangleq I_i e^{-l_i(\boldsymbol{\mu})}, \tag{3.1}$$

where $l_i(\boldsymbol{\mu})$ is the forward projection given by

$$l_i(\boldsymbol{\mu}) \triangleq \sum_j a_{ij}\mu_j. \tag{3.2}$$

Here, I_i is the mean number of counts in the absence of an attenuating medium, and μ_j is the linear attenuation coefficient in voxel j. The system matrix elements a_{ij} comprise the appropriately discretized point spread function relating the projection space to the image space; it is described further in Section 3.2. If projection i does not pass through voxel j, then a_{ij} is identically equal to zero.

3.1.3 Image reconstruction formulation

The basic goal of image reconstruction in transmission tomography is to estimate $\boldsymbol{\mu}$ from $\boldsymbol{d}$. In the context of statistical estimation problems, this is often achieved by maximizing the data likelihood objective function.

Maximum likelihood

Using a more sophisticated data model (in which the above model is a special case), O'Sullivan and Benac [66] derived an alternating minimization (AM) algorithm to find the ML solution. The problem was formulated as the double minimization of an $\mathcal{I}$-divergence over a linear and exponential family, thereby resulting in a closed-form update for each iteration. (Readers unfamiliar with $\mathcal{I}$-divergence are referred to the references cited within the above paper.) Omitting terms that do not depend on $\boldsymbol{\mu}$, the objective function for the mono-energetic case is

$$\mathcal{I}\left[\boldsymbol{d}||\boldsymbol{g}(\boldsymbol{\mu})\right] \triangleq \sum_i \left[-d_i \ln g_i(\boldsymbol{\mu}) + g_i(\boldsymbol{\mu})\right]. \tag{3.3}$$

It should be noted that for this simpler data model, minimization over the linear family is trivial, so there is no need to actually perform *alternating* minimization. We will still refer to our reconstruction method as the AM algorithm, however.

For the AM algorithm, we need to compute the two backprojections given by

$$\begin{aligned} b_j &\triangleq \sum_i a_{ij} d_i && (3.4) \\ \hat{b}_j^{(k)} &\triangleq \sum_i a_{ij} g_i(\hat{\boldsymbol{\mu}}^{(k)}), && (3.5) \end{aligned}$$

where $\hat{\boldsymbol{\mu}}^{(k)}$ is the estimate of $\boldsymbol{\mu}$ at iteration k. From O'Sullivan and Benac [66], the update

$$\hat{\mu}_j^{(k+1)} = \left[\hat{\mu}_j^{(k)} - \frac{1}{Z_j} \ln \left(\frac{b_j}{\hat{b}_j^{(k)}} \right) \right]_+ \tag{3.6}$$

monotonically decreases (3.3) (or equivalently, monotonically increases the Poisson likelihood objective function). The $(\cdot)_+$ notation is shorthand for $\max(\cdot\,, 0)$, and Z_j is a normalization function that satisfies $\sum_j a_{ij}/Z_j \leq 1 \ \forall i$.

For simplicity, Z_j can be chosen to be constant, in which case it should be the smallest constant such that the above condition on Z_j holds. Since a_{ij} is approximately equal to the length of projection i through voxel j, Z_j can be set equal to the maximum projection length through the reconstruction cylinder. Unlike the ICD method used by Thibault et al. [99], the AM algorithm does not need to be initialized with a FBP image; we start with a uniform image of all zeros.

Penalized likelihood

Since the measured data are noisy, it is necessary to regularize the optimization problem to prevent the algorithm from over-fitting the data through unrealistic images. With algorithms that perform simultaneous image updates on all voxels (like the AM algorithm), it is common to simply restrict the number of iterations, since low-frequency components tend to converge faster. Another standard regularization method is to add a penalty function to the original objective function as a means of discouraging large changes between neighboring voxels. If the reconstruction algorithm is run to convergence, the penalized objective function solely determines image quality – not the algorithm used to optimize the objective function [24].

We take an approach analogous to that of Erdoğan and Fessler [22] and other groups, and decouple the image variables of our penalized objective function in such a way that all the voxels can still be updated in parallel. Although the math is somewhat simpler for quadratic penalties, the derivation in Appendix A is general enough to allow for non-quadratic, edge-preserving functions.

To derive the penalized AM (P-AM) algorithm, we add a penalty term, $R(\boldsymbol{\mu})$, to the objective function used in the AM reconstruction, and weight it by a regularization parameter, λ. More specifically, the penalized $\mathcal{I}$-divergence objective function is

$$\Phi(\boldsymbol{\mu}) \triangleq \mathcal{I}\left[\boldsymbol{d}||\boldsymbol{g}(\boldsymbol{\mu})\right] + \lambda R(\boldsymbol{\mu}), \tag{3.7}$$

where

$$R(\boldsymbol{\mu}) \triangleq \sum_{j} \sum_{j' \in \mathcal{N}_j} w_{j,j'} \psi(\mu_j - \mu_{j'}). \tag{3.8}$$

For 3D regularization, we use the 26-voxel neighborhood $\mathcal{N}_j$ surrounding voxel j. The weights $w_{j,j'}$ control the relative contribution of each neighbor; we have used a standard choice of inverse distance between voxel centers. The potential function $\psi(t)$ is a symmetric convex function that penalizes the difference between the values of two neighboring voxels. This implies that rough regions of the image are penalized more heavily than smooth regions. The parameter λ is a scalar that reflects the amount of smoothing desired. A larger value will give emphasis to the penalty term (i.e., the prior expectation that the image will be smooth), whereas a smaller value will give more emphasis to the $\mathcal{I}$-divergence term (i.e., the discrepancy between the measured data and the data estimated by the model).

Due to its computational simplicity, we use a modified potential function from Lange [51],

$$\psi(t) \triangleq \frac{1}{\delta}\left[|\delta t| - \ln(1 + |\delta t|)\right], \tag{3.9}$$

where δ is a parameter that controls the transition between a quadratic region (for smaller t) and a linear region (for larger t). This penalty has also been studied by others, e.g., [27].

The addition of a penalty term prevents us from using a closed-form update at each iteration, as was possible in Section 3.1.3. Instead, we use Newton's Method on the decomposed $\mathcal{I}$-divergence and penalty terms, as shown in Appendix A. (If the OS method is used, λ is scaled down in each sub-iteration by the number of subsets used.)

3.2 Fully 3D system matrix

3.2.1 Derivation

To preserve the original resolution of the measured data, we modeled the native geometry of the scanner in our system matrix. The system matrix derivation, as first described in Politte and Whiting [72] for 2D fan-beam CT, and later expanded to 3D helical CT [73], leads to forward projection/backprojection operations that are neither ray-driven nor voxel-driven. Rather, the computations are based (more accurately) on the solid angle seen from the focal spot to a particular detector element, as well as the finite size and location of each voxel. In this regard, the method has somewhat similar characteristics to the distance-driven projector kernel proposed in De Man and Basu [17].

To calculate the system matrix elements for the helical CT geometry, first assume all detector elements and voxels are infinitesimally small, and that there is no focal spot blurring. Then the ideal 3D point spread function is a Dirac delta function of the three continuous spatial coordinates, and is denoted by $\tilde{h}(x, y, z|\beta, \gamma, \eta)$. It is nonzero only at points (x, y, z) along the source-detector ray specified by (β, γ, η). Integrating the ideal point-spread function over the spatial extent of voxel j therefore yields the length of intersection, $l_j(\beta, \gamma, \eta)$, between the ideal source-detector ray and the voxel, i.e.,

$$l_j(\beta, \gamma, \eta) \triangleq \underset{\text{voxel } j}{\int\int\int} \tilde{h}(x, y, z|\beta, \gamma, \eta)\, dx\, dy\, dz. \tag{3.10}$$

As derived in Politte and Whiting [72], and assuming isotropic propagation of photons from the x-ray source to detector array, a good approximation for the discrete point spread function is the average path length through a voxel along equiangular lines from the focal spot to a finite detector element. Mathematically, this is given by

$$a_{ij} \triangleq \frac{1}{\sigma_i} \cdot \frac{1}{\Delta\gamma} \int\limits_{\eta_{\min,i}}^{\eta_{\max,i}} \int\limits_{\gamma_{\min,i}}^{\gamma_{\max,i}} l_j(\beta_i, \gamma, \eta)\, d\gamma\ \cos\eta\, d\eta, \tag{3.11}$$

where β_i and the angular extent of the detector element are determined from the geometry of source-detector pair i. If blurring due to gantry rotation is to be modeled, $l_j(\beta, \gamma, \eta)$ would be averaged over the azimuthal extent of the measurement as well. The value $\Delta\gamma$ is the constant fan angle spacing of each detector element, and $\sigma_i \triangleq \int_{\eta_{\min,i}}^{\eta_{\max,i}} \cos\eta\, d\eta$. Since the integration in (3.11) is performed over a spherical surface, the proper spherical surface element has been used, i.e., $dS = (R_f + R_d)^2\, d\gamma \cos\eta\, d\eta$. (The radius factor cancels when taking the average.)

3.2.2 Fast computation using voxel traversal algorithm

To simplify the calculations, we replace the double-integral over the detector area with a 2D Riemann sum. We assume the detector elements are so small relative to the source-detector distance that the integral can be well-approximated by a grid of uniformly spaced points over the detector element.

Once this is done, computing a_{ij} is essentially a matter of computing the lengths of intersection between voxels and rays over the subdivided detector area. These intersection lengths can be calculated quickly using an algorithm that traverses a regular voxel grid along a given ray path [3]. Parallelization of these system matrix computations is trivial if desired.

3.2.3 Symmetry and storage details

Since the system matrix is extremely large, efforts were made to reduce the memory and computational requirements. This was accomplished by exploiting the symmetry

N_q	pitch = 0.5 ($z_{\text{feed}} = 12.0$ mm)	pitch = 1.0 ($z_{\text{feed}} = 24.0$ mm)	pitch = 1.5 ($z_{\text{feed}} = 36.0$ mm)
1	3.0	6.0	9.0
2	1.5	3.0	4.5
3	1.0	2.0	3.0
4	0.75	1.5	2.25
5	0.6	1.2	1.8
6	0.5	1.0	1.5
		$\vdots$	

Table 3.1: Table of valid reconstruction slice thicknesses (in mm) for various pitch settings using 16×1.5 mm collimation. These values are determined by $(z_{\text{feed}} \div 4)/N_q$, where N_q is the number of slices chosen to correspond to one quarter rotation.

in the helical CT geometry and by storing only the nonzero elements in a lookup table (LUT).

Most importantly, we have determined that only the source-detector pairs in the first quarter rotation of the gantry need to be computed and stored – regardless of the length of the scan. For this symmetry to be valid, an integer number of image slices must correspond to the distance the bed travels in a quarter rotation of the gantry. This is actually not much of a restriction, as any helical pitch may be used, and the reconstruction slice thickness can be made arbitrarily small. Table 3.1 gives an example for a 16-row scanner with 1.5 mm collimation at isocenter to illustrate the fact that this is a negligible limitation. In fact, it becomes even less limiting for scanners with larger axial coverage since they have a higher travel per rotation at a given pitch.

The symmetry is illustrated in Figure 3.2(a) for the case where the bed translates two slices per quarter-rotation (denoted by N_q). The solid box indicates the portion of the scan (i.e., the first quarter rotation) for which the LUT must be computed, while the dashed boxes represent the symmetrically-related LUTs. Also in this top part of the figure are two diagonal lines, which correspond to the axial coverage of the cone-beam at each view angle. As shown, the LUT stores the nonzero matrix elements for all the slices that are hit by the cone-beam projections in $\beta \in [0, \pi/2)$.

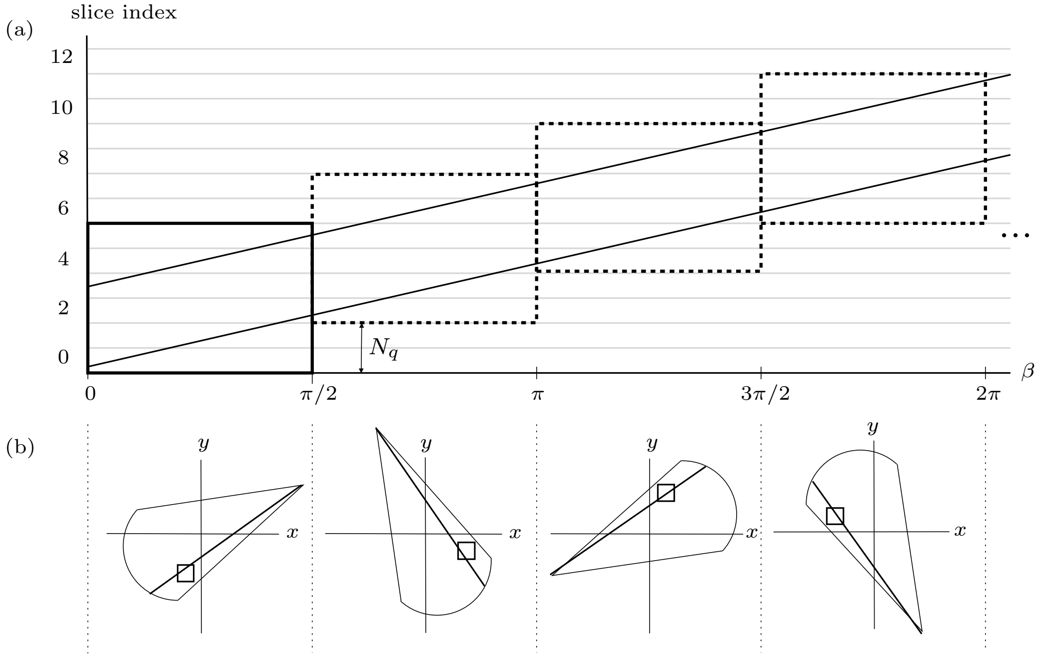


Figure 3.2: (a) Axial view of the quarter rotation symmetry found in helical CT. When an integer number of slices is chosen per quarter rotation of the gantry, only the system matrix elements in the first quarter rotation of the scan need to be computed and stored in the LUT (as indicated by dark solid box). (b) Transverse view of the quarter rotation symmetry. The length through the voxel shown in each 90°-rotated view is the same.

Figure 3.2(b) shows the familiar four-fold rotational symmetry in the x-y plane. This is used in conjunction with the appropriate slice offset to identify the symmetric voxels in each quarter rotation. Note that the top half of Figure 3.2 lines up vertically with the bottom half. Both the x and y dimensions must have the same (even) number of voxels, and the number of source angles per rotation must be divisible by four.

It should be noted that an additional two-fold savings can be obtained if the scanner does not employ quarter detector offset (QDO). Assuming the focal spot is at $z = 0$ when $\beta = 0$, the first measured source angle would need to be located at $\beta = \Delta\beta/2$, where $\Delta\beta$ is the source angle spacing. In this case, the LUT would only need to be computed and stored for half of the detector rows. We did not pursue eight-fold symmetry per rotation because the CT scanner used for testing does employ QDO. (The four-fold symmetry per rotation discussed above still holds even when QDO is used.) Appendix B provides a more mathematical discussion of the quarter rotation and axial mirror symmetries.

The precomputed LUT should fit in system memory to avoid having the reconstruction algorithm read in parts of the LUT from the hard disk during each iteration. We have developed an encoding of the sparse system matrix that requires three bytes per matrix element (plus a small header for the matrix). The encoding was designed to use the cache efficiently, and to work well with ordered subsets.

3.3 Parallelization scheme

This section discusses the method we developed to parallelize the AM and P-AM algorithms. The fact that the system matrix is symmetric for each quarter rotation makes it quite natural to implement parallelism at the granularity of a quarter rotation of data. Each processor or core $p \in \{0, 1, \ldots, N_p - 1\}$ is assigned a contiguous group of projections, $\mathcal{D}_p$, that is a multiple of a quarter rotation. With OS, each subset consists of evenly-distributed projections among all processors. For example, if two subsets are used, the even-indexed source angles on each processor will be used in the first sub-iteration, and the odd-indexed source angles will be used in the second sub-iteration.

Regardless of whether OS is used, this design allows for theoretically perfect load balancing (in the absence of communication overhead) during forward and backprojections, since each processor will make use of exactly the same number of nonzero a_{ij} elements. In fact, even the communication among processors is small for typical problem sizes. This is due to the inherent locality exploited by our implementation; each quarter rotation of data is related to a small local neighborhood of slices, as already shown in Figure 3.2(a).

Our approach is currently most appropriate for a shared memory system (such as a multi-processor, multi-core workstation), so it is necessary to distinguish whether each array is stored in shared memory or private memory. The only arrays that need to be stored in shared memory are $\hat{\boldsymbol{\mu}}^{(k)}$, the two backprojection images $\boldsymbol{b}$ and $\hat{\boldsymbol{b}}^{(k)}$, and the system matrix LUT (due to its large size and the fact that each processor needs read-only access to the entire LUT). All other arrays are stored in private memory. The image arrays have been stored in shared memory since a given slice may need to be accessed by multiple processors during each iteration.

Figure 3.3 illustrates the main components of the parallelized algorithm. The vast majority of this flowchart is enclosed in a dashed box, which indicates that these steps are executed in parallel.

3.3.1 Reading in the data

Step 1 of the flowchart simply reads in the assigned portion of the measured data to the private memory associated with each processor.

3.3.2 Forward and backprojections

Step 2 contains the forward and backprojection calculations. This step takes up almost all of the time spent in each iteration. The forward projections and exponentiations are straightforward, because this part only requires read access from the shared $\hat{\boldsymbol{\mu}}^{(k)}$ image. However, if we were to perform the backprojections directly into the shared backprojection images, we would have serious memory contention issues since multiple processors would be writing to the same array elements simultaneously. Instead, we perform partial backprojections to smaller private image arrays associated with each processor. This eliminates any need for synchronization during the partial backprojection computations.

It is easiest to illustrate this concept with an example. Referring to Figure 3.2(a), suppose there are two processors; the first one is assigned $\beta \in [0, \pi)$, and the second one is assigned $\beta \in [\pi, 2\pi)$. It can be observed that processor 0 will only ever need to access slices 0-7, while processor 1 will only ever need to access slices 4-11. Therefore, each private backprojection array consists of eight slices, and each processor can easily determine what its starting slice index should be.

Although the backprojection of the measured data is constant from iteration to iteration, when OS is used, a full backprojection image corresponding to each subset is

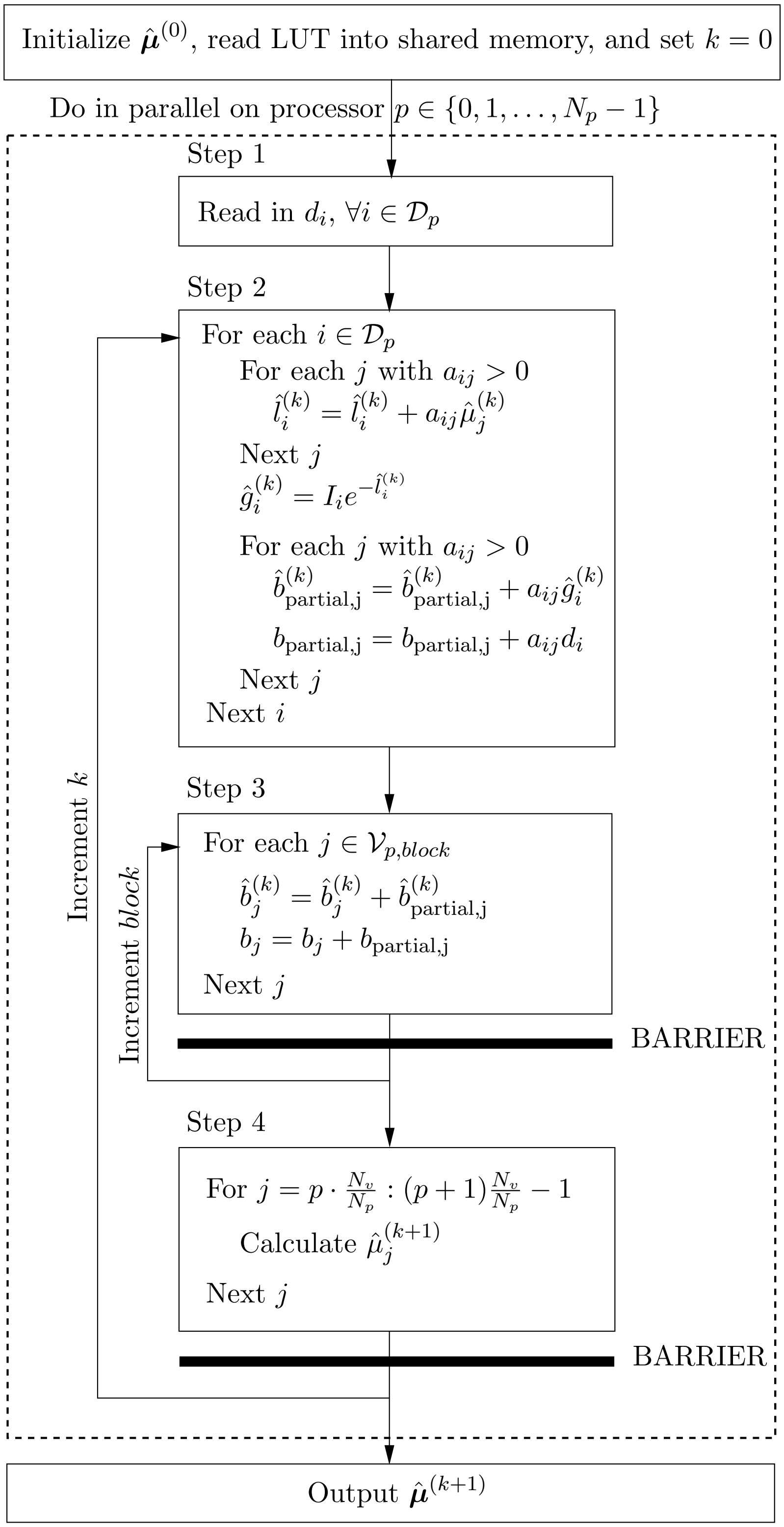


Figure 3.3: Flowchart for the parallelized AM and P-AM algorithms.

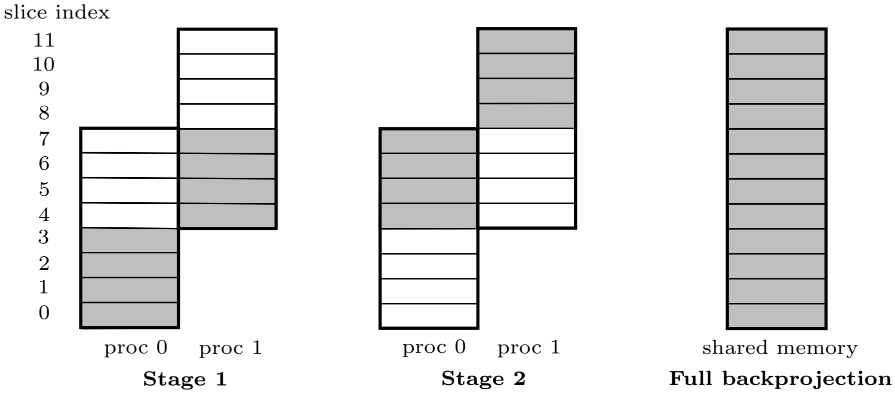


Figure 3.4: Summing of private partial backprojections on processors 0 and 1 into complete backprojection image. At each stage, the shaded block of slices from each processor are simultaneously summed into the full backprojection image. In this example, both processors use slices 4-7, so this approach systematically adds these slices with minimal synchronization overhead between the stages. The code determines the maximum block size that can be summed concurrently by all processors.

required. When using a large number of subsets, it is infeasible to store all of these 3D images in memory. Therefore, we compute the measured data backprojection in each (sub-)iteration, along with the estimated data backprojection.

3.3.3 Summing the partial backprojections

Step 3 takes these partial backprojections, and systematically sums them into the full backprojection images, $\boldsymbol{b}$ and $\hat{\boldsymbol{b}}^{(k)}$. Figure 3.4 illustrates the process by which independent slices from each partial backprojection (denoted $\mathcal{V}_{p,block}$ in the flowchart) can be added simultaneously without memory contention. After each block, a synchronization construct called a barrier is used to ensure each processor has finished summing the current block of slices to the complete backprojection arrays.

3.3.4 Image update

Finally, Step 4 performs the image update. N_v is the number of voxels in the entire image. Since the image variables were decoupled in the AM and P-AM algorithms, each processor simultaneously performs the update on a unique group of slices. This is followed by another barrier before proceeding to the next iteration or stopping.

Using the regularized AM method, there is one slight complication that takes place during the calculation of the penalty derivative term in (A.9). Using 3D regularization, penalties are calculated using neighboring slices. Therefore, when a processor calculates the penalty for either its first or last slice, the calculation involves the last or first slice, respectively, on a neighboring processor. Suppose the updates are done starting from the smallest slice index up to the largest slice index on a given processor. Then the last (yet to be updated) slice will use the next processor's first slice, which *will* have already been updated. To avoid the unknown effects of using both updated and non-updated slices in the penalty calculation, we store the not-yet-updated first slice on each processor in temporary memory for use in the penalty calculation of the last slice on other processors. (Although there is a remote possibility that a processor would only finish computing the penalty on its first slice after the neighboring processor's last slice had already been updated, we did not consider this unlikely scenario in our current implementation.)

3.4 Experiments

We implemented our algorithm using OpenMP, an industry-standard parallel computing library designed for shared memory systems. The C code was compiled using the Intel Compiler 9.1 with certain optimizations enabled. The code was run on a Dell Precision T7400 workstation, equipped with two quad-core Intel Xeon X5450 processors (3.0 GHz, 1333 MHz front-side bus) and 16 GB RAM (667 MHz). The operating system running on this machine was Microsoft Windows XP Professional x64.

Phantom and clinical data were acquired on a Siemens Somatom Sensation 16 scanner (Siemens Medical Solutions, Forchheim, Germany) without using the flying focal spot mode. The scanner acquires 1160 views per rotation, using a 16 row × 672 channel curved detector array. The distance between the source and isocenter is 570 mm, and the distance between the source and detector is 1040 mm. Our phantom experiments used the Catphan 500 (The Phantom Laboratory, Salem, NY).

We first scanned the high resolution module of the Catphan phantom, which contains a ring of resolution bar patterns ranging from 1-21 line pairs per cm. This module also contains a tungsten carbide bead of diameter 0.28 mm that is positioned 20 mm below the x axis. The data from this scan was reconstructed using three algorithms: (1) P-AM with 145 subsets, (2) our implementation of the most basic helical FDK method described in Stierstorfer et al. [93], and (3) investigational offline reconstruction software provided by Siemens Medical Solutions that is similar to the software on the scanner console [80]. Convergence in our iterative algorithm was defined as being met when the mean absolute value change in voxel values in each slice was less than 2.0×10^{-5} mm^{-1} ($\sim$1 HU) from one iteration to the next. Captions on images in the next section indicate the iteration number at which this criterion was satisfied.

The bar pattern reconstructions will be shown for visual comparison at matched noise level (as determined by the standard deviation in a central constant-attenuation region of interest), while reconstructions of the bead will be used to generate a resolution-noise tradeoff curve both in-plane and axially. For the P-AM algorithm, we fix δ at 0.0002 mm^{-1} (chosen to match the equivalent parameter, c, in Thibault et al. [99]) and sweep λ over a wide range of values to produce the tradeoff curve. For both the helical FDK and Siemens algorithms, we use the body kernels B10-B60 described in Schaller et al. [83]. Voxel size was $0.5 \times 0.5 \times 0.75$ mm for all three reconstructions.

Next, we will show the effect of using ordered subsets on the convergence rate of the AM algorithm. Using the data from a clinical abdominal scan, we performed a reconstruction using AM without OS, and AM with 145 subsets. (The maximum number of subsets allowed by our implementation is 290, which corresponds to subsets containing one view per quarter rotation; this was deemed to be too aggressive, as it sometimes resulted in image artifacts.) We also examined the effect on convergence rate of initializing the AM iterations with the helical FDK image. In all the other results presented here, we initialized the iterations with an image of all zeros.

Data from the abdominal scan and a hip implant scan were reconstructed using AM and P-AM, respectively (with 145 subsets) and the Siemens reconstruction software. The abdominal reconstruction was done using 16 gantry rotations of data collected with pitch = 1.0 and 16×1.5 mm collimation at isocenter. The reconstructed image was 512 × 512 × 224 with 1.0 × 1.0 × 2.0 mm voxels. For the hip implant reconstruction, we used 16 gantry rotations of data collected at pitch = 0.5 and 16×0.75 mm collimation. The reconstructed image was 512 × 512 × 160 with 1.0 × 1.0 × 0.75 mm voxels. Efforts were made to match the noise in one relatively uniform region of interest in the P-AM and Siemens reconstructions.

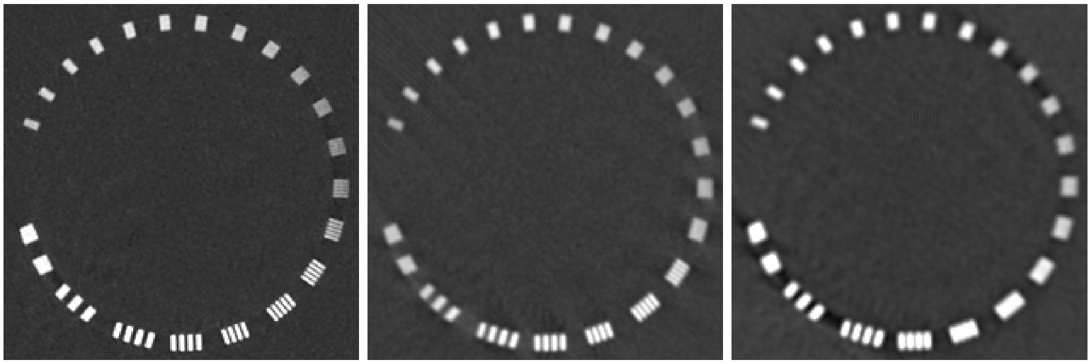
Finally, we performed timing tests using the full-scale abdominal dataset to investigate the performance of our computational approach on clinically-sized data. The reconstruction was done using one iteration of AM without subsets, and one iteration of AM with 145 subsets. In the latter case, the image update step is invoked 145 times, which could have an impact on timing performance due to the need for more frequent synchronization.

We tested the timing of both regularized AM and unregularized AM, but found negligible difference; hence, we will only show results for the unregularized algorithm. The timing results are based on Steps 2-4 from Figure 3.3, which prevents file I/O operations from being included (e.g., initially reading in the LUT). To generate a speedup curve, a separate serial version of the implementation was written and compiled without OpenMP. Step 3 was omitted from the serial code since the partial backprojection arrays are equivalent to the complete backprojection arrays in that case.

3.5 Results

3.5.1 Resolution phantom

In Figure 3.5, reconstructed images of the line pair test pattern are shown – all reconstructed using the same dataset. The phantom was not aligned perfectly with the scanner, and the voxels were not small enough to be able to resolve the very high frequency line pair spacings. However, these images are at matched noise level in the center of the image, and clearly show the better resolution recovery of the P-AM algorithm.



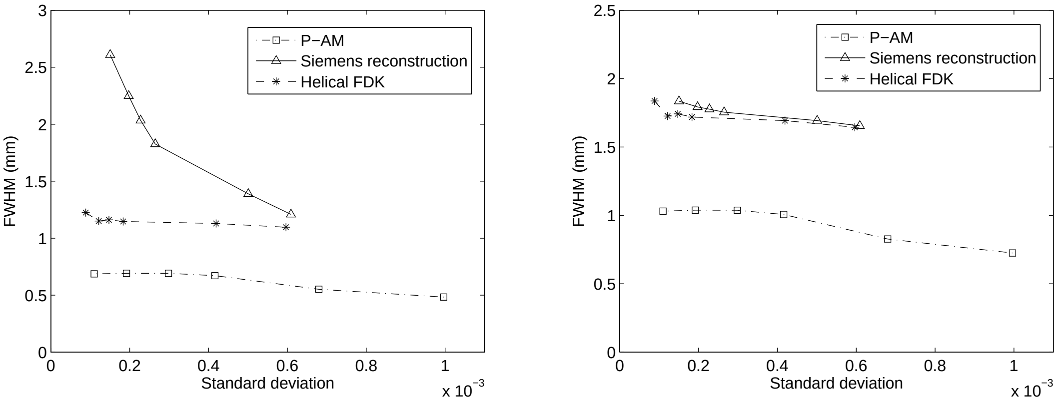
(a) (b) (c)

Figure 3.5: Line pair phantom reconstructions at matched noise. Window width=0.03 mm^{-1}, center=0.03 mm^{-1}. Voxel size $0.5 \times 0.5 \times 0.75$ mm. Scan parameters: 200 mAs, pitch 1.0, 16×0.75 mm collimation. (a) P-AM (iteration 9, $\lambda = 3000$); (b) Siemens reconstruction (filter B40); (c) helical FDK (filter B45).

The resolution-noise tradeoff curve is shown in Figure 3.6 for the transverse and axial directions. The transverse full width at half maximum (FWHM) resolution was obtained by fitting a Gaussian curve to the horizontal profile of maximum intensity. This was repeated for the axial direction in a similar manner. The P-AM curve gives the best quality with respect to this tradeoff because it is the farthest down on the plots and is relatively flat. These curves are specific to the bead location and the choice of reconstruction parameters. All three reconstruction methods used the same voxel size and spacing. Although many other filter designs could have been tested with the analytical methods, a more extensive quantitative image analysis is beyond the scope of this work.

3.5.2 Ordered subsets and FDK initialization

The use of OS has a significant effect on convergence rate. Figure 3.7 shows that using a large number of subsets leads to much faster convergence than not using OS. Although the AM objective function is not guaranteed to converge monotonically when using OS, we have yet to see an example using $\leq$145 subsets where it has not decreased monotonically at each iteration. The reconstructions we have performed using 145 subsets tend to converge after about 10-15 iterations. Increasing λ in the



(a) (b)

Figure 3.6: Resolution-noise tradeoff curves for 0.28 mm diameter tungsten carbide bead embedded in uniform background in Catphan phantom. Voxel size 0.5×0.5×0.75 mm. Scan parameters: 200 mAs, pitch 1.0, 16×0.75 mm collimation. (a) transverse direction; (b) axial direction.

P-AM algorithm (for a given dataset) usually leads to faster convergence as well, due to the image-based convergence criterion.

This plot also shows the effect of initializing the AM iterations with the helical FDK image. This does lead to fewer iterations to reach convergence, but its effect is not very significant for 145 subsets. We have observed that if our algorithm is initialized with a very noisy FDK image, the noise remains in the AM image even after many iterations.

3.5.3 Clinical datasets

We show the axial, coronal, and sagittal views of the abdominal reconstructions in Figure 3.8. The AM and Siemens reconstructions are quite similar for this dataset, which provides further validation of our method.

Figure 3.9 shows the hip implant reconstructions. The noise structure is very different between these two reconstructions. In the Siemens reconstruction, the excessive noise due to the metal implant is propagated throughout the FOV, whereas in the P-AM

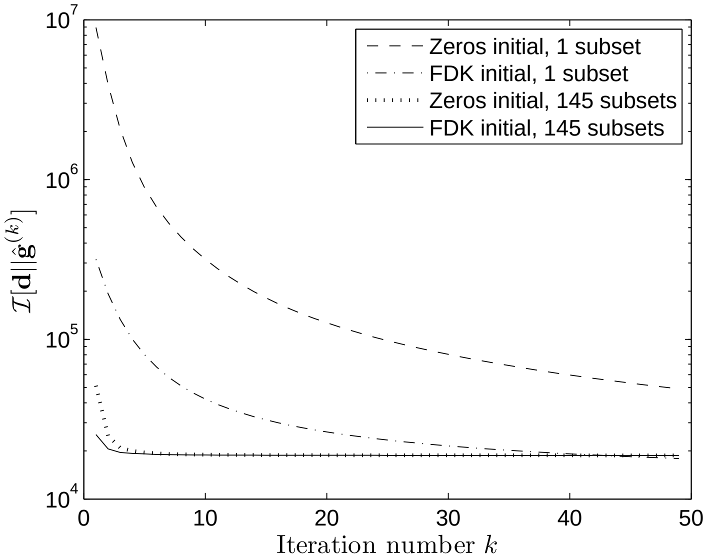


Figure 3.7: Unregularized objective function versus iteration comparing the use of OS and FDK initialization of the AM algorithm. Reconstructions of the abdominal dataset were used to generate this plot.

image, it is not. However, both images suffer from beam hardening artifacts, as evidenced by the shadows.

3.5.4 Timing Performance

As mentioned above, the abdominal dataset was used as a benchmark for the timing performance. The wall-clock time to run one iteration using all eight cores was 16.1 minutes with 145 subsets, and 14.1 minutes without subsets. Using the Intel Thread Profiler, we determined that in the case of no subsets, 96.2% of the execution time was in parallel, with the rest of the time spent waiting in barriers. Using 145 subsets, 95.1% of the execution time was in parallel, with the rest in barriers. These profiler results confirm that load balancing was well-maintained within each iteration.

The speedup as a function of the number of cores is plotted in Figure 3.10. As expected, the 145 subsets speedup is less than the speedup without OS, due to more synchronization points in each iteration. The plot also indicates superlinear speedup, i.e., speedup beyond the theoretical algorithmic capability. This type of behavior is

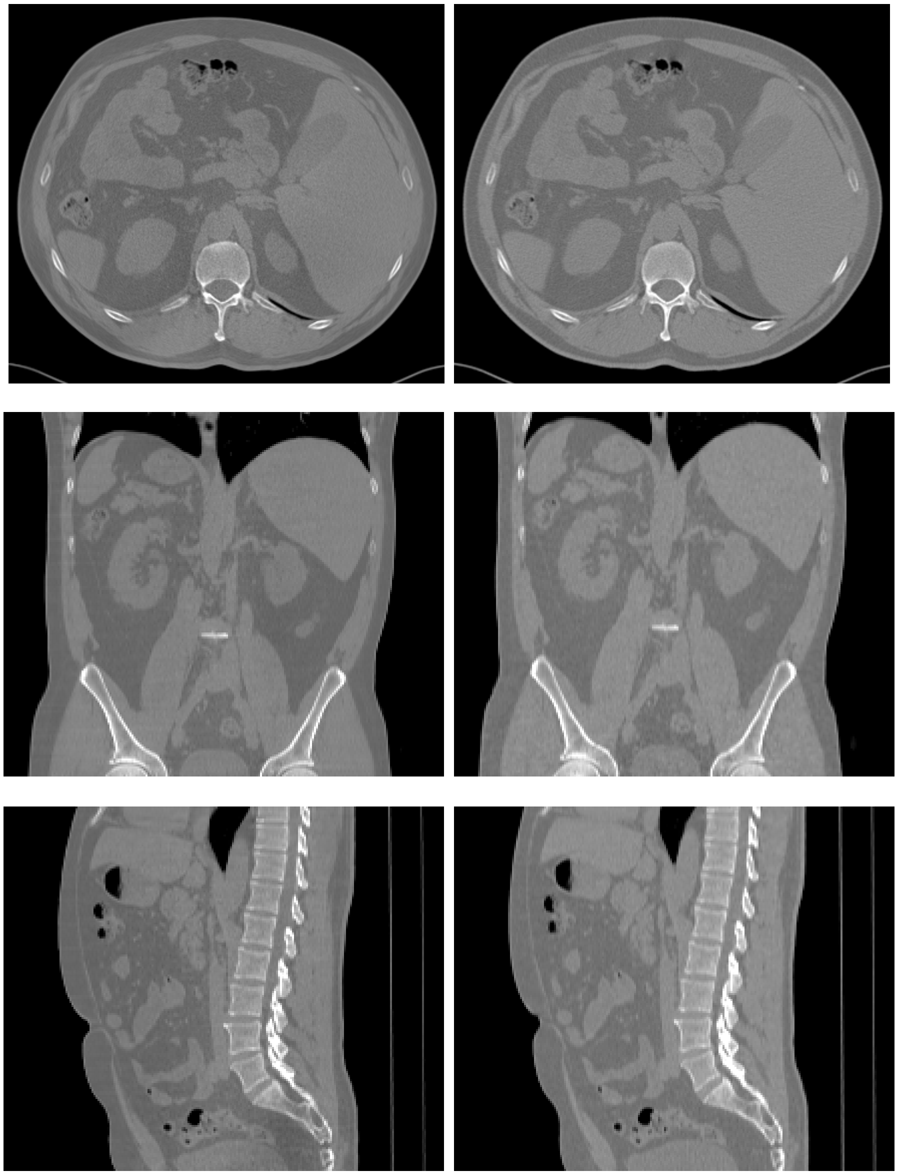
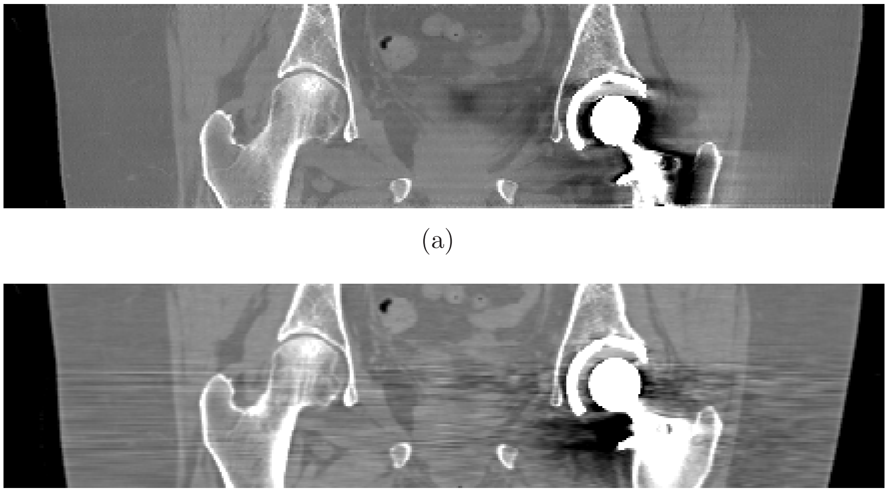


Figure 3.8: Clinical abdominal reconstructions at matched noise. Window width=0.03 mm^{-1}, center=0.02 mm^{-1}. Voxel size $1.0 \times 1.0 \times 2.0$ mm. Scan parameters: 180 mAs, pitch 1.0, 16×1.5 mm collimation. Rows correspond to axial, coronal, and sagittal views. Left column: AM (iteration 9); right column: Siemens reconstruction (filter B45).



(a)

(b)

Figure 3.9: Clinical hip implant reconstructions at matched noise (in a slice not affected by metal implant). Window width=0.03 mm^{-1}, center=0.02 mm^{-1}. Voxel size $1.0 \times 1.0 \times 0.75$ mm. Scan parameters: 450 mAs, pitch 0.5, 16×0.75 mm collimation. (a) P-AM (iteration 14, λ=6300); (b) Siemens reconstruction (filter B20).

fairly common for memory-intensive applications that are run on multi-core systems and some other architectures, and is largely due to the performance of the processor cache [98].

In our application, a smaller amount of data is assigned to each core as the number of cores increases. This leads to a larger effective cache size available to each core (i.e., a larger percentage of the data can fit in the fast L2 cache instead of remaining in RAM). Through the use of the Intel VTune Performance Analyzer, we found that the L2 cache miss rate approximately doubled with half as many cores. These speedup results were not only repeatable on the same machine, but similar results were also observed using the same parallelized implementation on a large-scale shared memory system, the SGI Altix 3700, at the National Center for Supercomputing Applications (University of Illinois at Urbana-Champaign). However, it should be noted that the speedup will depend to some extent on the size of the dataset and the scan parameters. The performance shown here is for one typical clinical scan protocol.

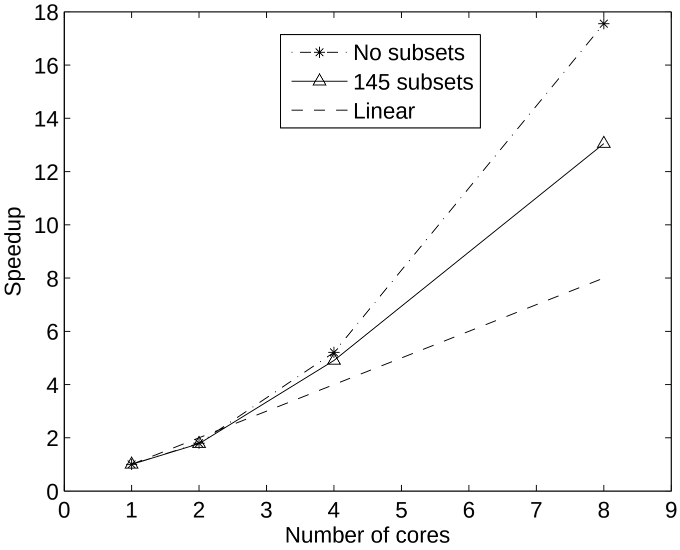


Figure 3.10: Speedup plot for one AM iteration with and without OS. Linear speedup also shown. Performance based on reconstruction of clinical abdominal dataset. Size of data space: 18 560 view angles, 672 channels per detector row, 16 detector rows. Size of image space: 512 × 512 × 224 voxels.

3.6 Discussion

The images and other results presented show some of the benefits of using statistical reconstruction for helical CT. We demonstrated the improvement in the resolution-noise tradeoff in both the transverse and axial directions, and showed that OS provides the means to accelerate the reconstruction considerably. A further benefit of our algorithm is that initialization with an FDK image is not necessary; this simplifies the reconstruction workflow. More thorough image quality evaluation would certainly be needed before deployment in the clinical environment, but these initial results indicate it is worth pursuing further.

Our current implementation is limited mainly by the memory needed to hold the LUT. The LUT size changes with the voxel dimensions, the scan parameters, and the number of detector element subdivisions (currently 20×20), but for typical parameters on a 16-row scanner, the LUT takes about 10-13 GB in our sparse encoded format. If memory is an issue, the system matrix could alternatively be computed on-the-fly.

In this case, only a small number of detector element subdivisions would be possible using current workstation capabilities (in order to keep the reconstruction runtime to a minimum). All the concepts from Sections 3.2 and 3.3 would still apply, regardless of whether the LUT is stored in memory or computed on-the-fly.

A significant additional computational savings could come from precomputing the measured data backprojection image (averaged over all subsets), as was done in Erdoğan and Fessler [22]. This would reduce the computations per iteration to one forward projection and one backprojection (instead of one forward projection and two backprojections).

We did not test our parallelized algorithm on more than one node. It would be possible to extend it to a multi-node system using a hybrid OpenMP/MPI approach, or by using some other parallelization library like Intel Cluster OpenMP. Each node would need a copy of the LUT, as well as the portion of the image that affects any projections on that node. The favorable timing performance results imply that our implementation can scale well to more than one node, although the increase in speedup with additional processors would likely drop due to the need for inter-node communication.

Overall, we have presented the implementation details for a new fully 3D statistical iterative reconstruction algorithm for multislice helical CT. Few statistical algorithms and results have appeared in the literature pertaining to this scan geometry, despite its widespread clinical usage. To reduce the computational burden of reconstructing large clinical datasets iteratively, we have taken advantage of the quarter rotation symmetry that applies to helical CT, and parallelized the algorithm according to this symmetry. The image quality results obtained with this new method demonstrate its potential to perform well in comparison to existing analytical methods.

Chapter 4

Reconstruction of Incomplete CT Datasets

Motivated by the promising 2D results attained with the incomplete data techniques of Snyder et al. [89], we present experimental results showing the effectiveness of using these methods to reduce such problems in fully 3D statistical reconstructions from multislice helical CT data. We will use a modified version of the reconstruction implementation described in Chapter 3 for the experiments in this chapter.

The rest of the chapter is organized as follows. Our application of Snyder et al. [89] to fully 3D reconstructions is given in Section 4.1. Sections 4.2 and 4.3 describe our initial experiments on simulated phantom data and clinical data. Following this, we provide a discussion of the results in Section 4.4.

4.1 Theory

In our approach, we update all the voxels in the image, with the potential to accurately reconstruct even the values outside the FOV. We use Snyder et al. [89] as the basis for our present work. For completeness, we give an overview of the two methods contained in that paper using its original notation. We take $\mathcal{X}$ to represent an image space that fully contains any part of the object affecting the measured rays. The

---

Some material in Chapter 4 is based on “Missing Data Estimation for Fully 3D Spiral CT Image Reconstruction,” by D. B. Keesing, J. A. O’Sullivan, D. G. Politte, B. R. Whiting, and D. L. Snyder, which appeared in Proceedings of SPIE, Medical Imaging 2007: Physics of Medical Imaging, Vol. 6510; © 2007 SPIE.

measured rays comprise the incomplete data space denoted by $\mathcal{D}_{\text{inc}}$, while $\mathcal{D}$ is the data space that corresponds to a complete set of projections through all elements of $\mathcal{X}$. The missing measurements are defined to be $\mathcal{D}_{\text{miss}} \triangleq \{i \in \mathcal{D} \,|\, i \notin \mathcal{D}_{\text{inc}}\}$.

O'Sullivan and Benac [66] showed that maximization of the transmission Poisson likelihood function is equivalent to double-minimization of the $\mathcal{I}$-divergence between elements of a particular linear and exponential family. This can be thought of as minimizing the discrepancy between the measured data and the mean estimate of the data. Snyder et al. [89] modified the linear family over which to perform the minimization to account for the fact that $\mathcal{D} \neq \mathcal{D}_{\text{inc}}$ in missing data problems. Using a standard monoenergetic, scatter-free Poisson data model in the derivation, the image update step

$$\hat{\mu}_j^{(k+1)} = \max\left\{\hat{\mu}_j^{(k)} - \frac{1}{Z_j}\ln\left[\frac{\sum_{i\in\mathcal{D}_{\text{inc}}} a_{ij}\, d_i + \sum_{i\in\mathcal{D}_{\text{miss}}} a_{ij}\, \hat{g}_i^{(k)}}{\sum_{i\in\mathcal{D}_{\text{inc}}} a_{ij}\, \hat{g}_i^{(k)} + \sum_{i\in\mathcal{D}_{\text{miss}}} a_{ij}\, \hat{g}_i^{(k)}}\right],\ 0\right\} \tag{4.1}$$

guarantees monotonic convergence to the minimum $\mathcal{I}$-divergence, where $\hat{\mu}_j^{(k)}$ is the attenuation value at the kth iteration at voxel j, Z_j is a normalizing function, a_{ij} is the system matrix relating the image and data spaces, d_i is the data collected at source-detector pair i, and $\hat{g}_i^{(k)}$ is the data estimate at source-detector pair i based on the kth image estimate. As in Chapter 3, $\hat{g}_i^{(k)} \triangleq I_i \exp\left[-\sum_j a_{ij}\hat{\mu}_j^{(k)}\right]$, where I_i is the incident flux detected by source-detector pair i in the absence of attenuation. It is important to note that the update step is applied to all $x \in \mathcal{X}$. Equation (4.1), as shown, is referred to as Method 2; setting $\mathcal{D}_{\text{miss}} = \emptyset$ yields Method 1. Therefore, Method 1 uses just the measured data (even if incomplete), whereas Method 2 estimates the missing data at each iteration in conjunction with the update of all elements in the image space. In the case where $\mathcal{D} = \mathcal{D}_{\text{inc}}$, we obtain the original AM algorithm [66] (for the simplified data model used here).

4.2 Experiments

We first generated a complete sinogram of the inner slices of the 3D NCAT phantom. To effectively reduce the radius of the FOV, we then artificially truncated 60 detector elements in all detector rows (30 from each end). In slices with smaller body cross-sections, this truncation did not affect the nonzero projections; however, projections were truncated for a number of slices containing wider cross-sections. This experiment therefore allowed us to evaluate the performance of different reconstruction techniques on a simulated dataset in which the patient was larger than the scanner FOV. Since the full data was also available, it was straightforward to compare the accuracy with and without the missing data approaches.

Since the beginning and end slices of the reconstruction volume are severely undersampled, we initialized these outer slices to the values of the nearest fully sampled slice after the first iteration of the algorithm. The rationale is that the undersampled slices at the ends of the reconstruction cylinder will typically be similar to the nearest fully sampled slice. This new initial guess, which is more representative of the attenuation in the long object than an image of all zeros, should help the algorithm reach convergence in fewer iterations. An alternative choice for the starting guess could be a filtered backprojection (FBP) image.

By definition, $\mathcal{D}_{\text{miss}}$ consists of all the rays that are not measured but that fit completely within the reconstruction volume boundary defined by the start and end slice. In our implementation, we took $\mathcal{D}_{\text{miss}}$ to be all the rays along the helical source trajectory (beyond $\mathcal{D}_{\text{inc}}$) whose affected voxels were contained entirely within the reconstruction cylinder. Virtual extension of the helical source trajectory (i.e., estimation of the missing rays in Method 2) could therefore employ the same pre-computed system matrix lookup table that was used in the computations for $i \in \mathcal{D}_{\text{inc}}$.

For all of our experiments below except the clinical data experiments, we generated numerically simulated data for the NCAT phantom [85] using a system matrix twice as fine as that which was used in the reconstructions. Reconstructions using Method 1 or Method 2 were performed for noiseless and noisy datasets, starting with an initial image of all zeros. The noisy datasets were obtained by running the simulated mean data through a Poisson random number generator. Transverse and longitudinal truncation were studied separately.

4.3 Results

4.3.1 Transverse Truncation

In our simulation experiments, we compare the methods described in this chapter by reconstructing images of the NCAT phantom on a 128×128×84 voxel grid. All of the images in Figure 4.1 are shown after 100 full iterations using 73 ordered subsets. Decrease in the $\mathcal{I}$-divergence from iteration-to-iteration had become very small well before 100 iterations (see Figure 4.2), although this is not to say the image had fully converged; small improvements to the image were still occurring beyond 100 iterations.

Without truncation, there were 168 detector elements per row, 8 detector rows, and 584 views per rotation. A pitch of 2 was used in the simulations. When artificial truncation was applied, only the measured data corresponding to the central 108 detector elements per row was available during the reconstruction. The truncated FOV circle can be readily seen in some of the images. The smallest square that encompassed the reconstruction circle subtended by 108 detector elements was 84×84. When using Method 2, the missing data was estimated for all 60 missing detector elements per row.

In each panel of Figure 4.1, slice 70 is shown on the left, and slice 74 is shown on the right. Figure 4.1(a) shows the truth image; the gray background value was approximately 0.04 mm^{-1}. Without using a missing data approach to reconstruct this truncated dataset, the reconstruction would by default take place on an 84×84×84 image. The results of such a reconstruction are shown in Figure 4.1(b). Note the bright voxel values at the boundary of the FOV; no image values are reconstructed outside the FOV in this case. Figures 4.1(c) and 4.1(d) show the results of applying Methods 1 and 2, respectively, to noiseless data. Although the methods behave differently, they are clearly able to perform a good reconstruction within the FOV and a reasonable reconstruction outside the FOV.

To illustrate the convergence properties of the above methods, Figure 4.2 shows the $\mathcal{I}$-divergence as a function of iteration number for the corresponding reconstructions in Figures 4.1(b)-4.1(d). For comparison, the "complete data available" curve shows

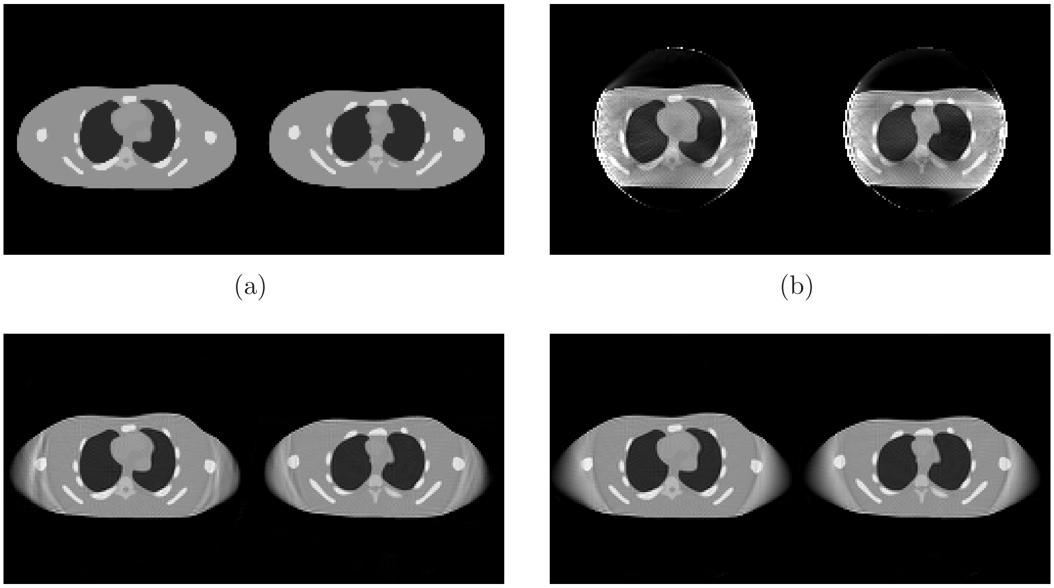

(a) (b) (c) (d)

Figure 4.1: Truth image and unregularized image reconstruction results shown for slices 70 and 74 using various methods. Image window set to [0.0, 0.07] mm^{-1}. (a) Truth image; (b) reconstruction without a missing data approach; (c) Method 1 reconstruction using noiseless data; (d) Method 2 reconstruction using noiseless data.

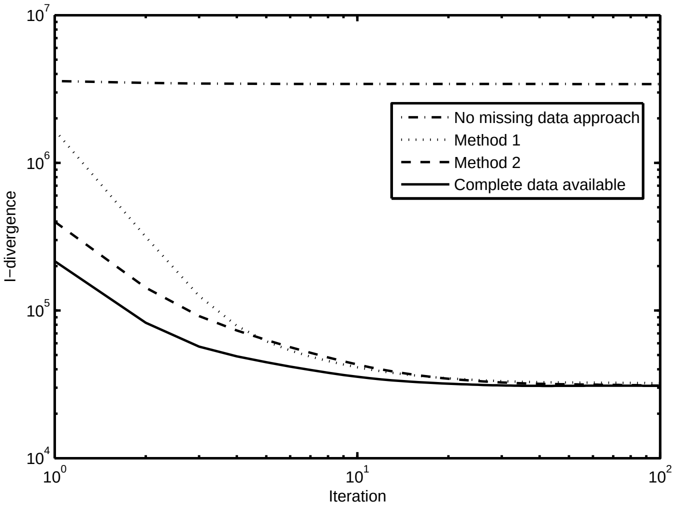
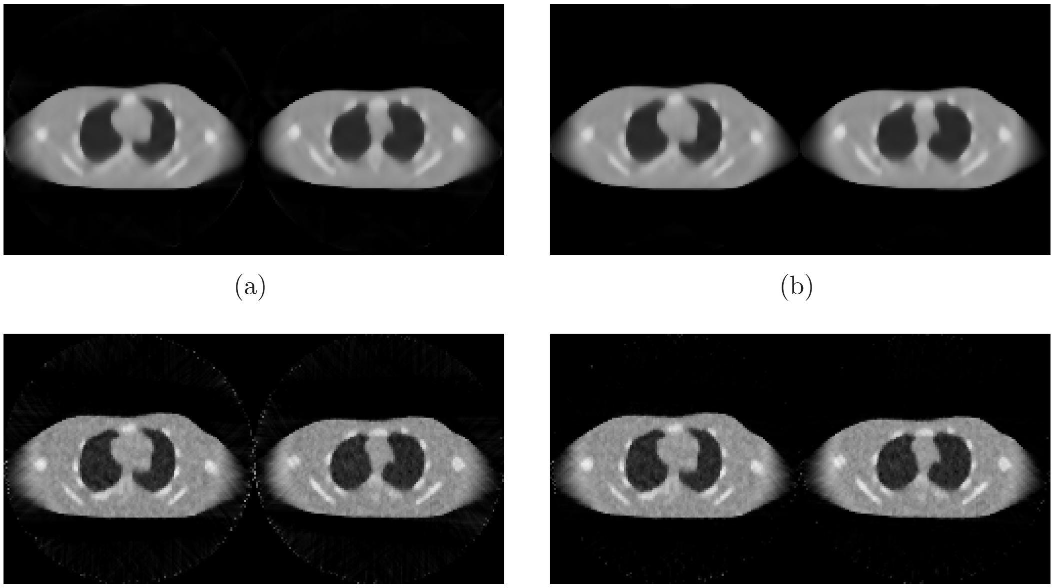


Figure 4.2: $\mathcal{I}$-divergence versus iteration number for the unregularized reconstructions. The "complete data available" curve corresponds to a reconstruction without missing data, and the other three curves correspond to Figures 4.1(b)-4.1(d), respectively.



(a) (b) (c) (d)

Figure 4.3: Regularized image reconstruction results shown for slices 70 and 74 using various methods. Image window set to [0.0, 0.07] mm^{-1}. (a) Method 1 reconstruction using noiseless data; (b) Method 2 reconstruction using noiseless data; (c) Method 1 reconstruction using noisy data; (d) Method 2 reconstruction using noisy data.

the convergence behavior had all the data been available for reconstruction. With the exception of the "no missing data approach" curve, the methods converge toward the same $\mathcal{I}$-divergence value.

As seen in Figures 4.3(a) and 4.3(b), the delineation between the truncated FOV and extended FOV is less noticeable when using regularization. These two images were generated specifically to illustrate this point; however, the images are excessively smooth compared to the unregularized images. Figures 4.3(c) and 4.3(d) show the results of running Methods 1 and 2, respectively, on noisy data. The mean incident photon count per detector element in air was 500. (We were able to use such a low count because the object volume was so small and no high density attenuators were present in the object.) We used the log cosh penalty [30] with $\beta = 1$ and $\delta = 100$ for all the image reconstructions shown in Figure 4.3.

We also applied Methods 1 and 2 to real clinical data collected on a Somatom Sensation 16, a 16-row spiral CT scanner (Siemens Medical Systems, Forchheim, Germany).

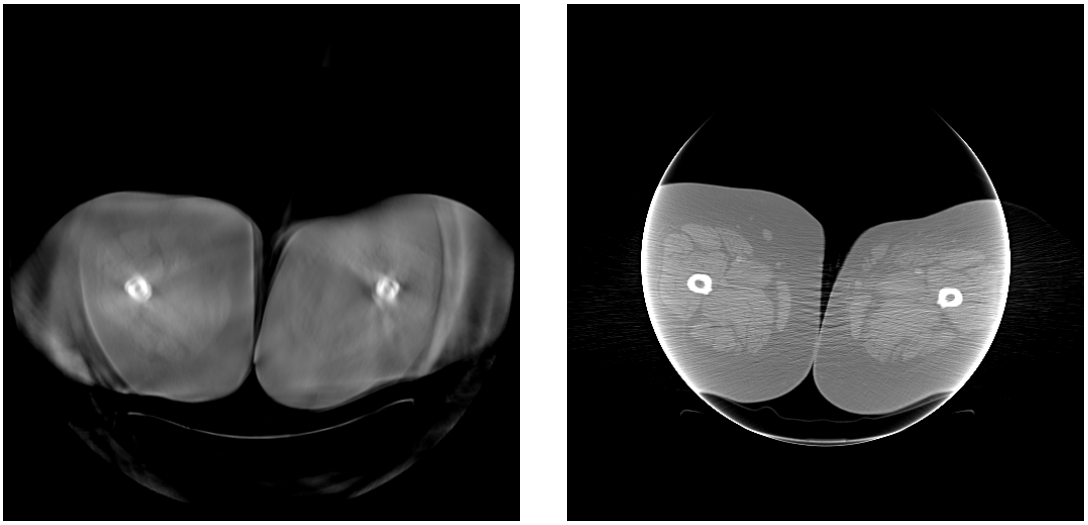

(a) (b)

Figure 4.4: One slice of obese patient reconstruction. Image window set to [0.005, 0.035] mm^{-1}. (a) Method 1 after six iterations; (b) FBP reconstruction.

This scanner has a FOV of 50 cm, and a bore diameter of 70 cm. The first clinical dataset was of a large patient, and therefore had a significant amount of transverse truncation. Figure 4.4(a) shows a slice of the reconstructed image (512×512×64) after performing six full iterations of Method 1 (using 145 ordered subsets). For comparison, we show the corresponding FBP image in Figure 4.4(b). The Siemens FBP software automatically stores its reconstructed images in HU, so we converted it to an image of attenuation coefficients (assuming an effective energy of $\sim$75 keV) in order to display the two images using the same window setting. Despite the same display window, the two images look rather different. FBP was successful in reconstructing the detailed structure inside the scanner FOV, but not in the extended FOV. Our method does not show such fine structure, but the full extent of the patient in this slice is visible. Even though our reconstruction could be improved, applications such as attenuation correction might perform better using this reconstructed image because the support of the patient is more accurately quantified. The lack of high contrast is not surprising given the small number of iterations performed.

The image reconstructions (512×512×176) from a second clinical dataset are shown in Figure 4.5. Scan parameters were typical for abdominal imaging. We artificially

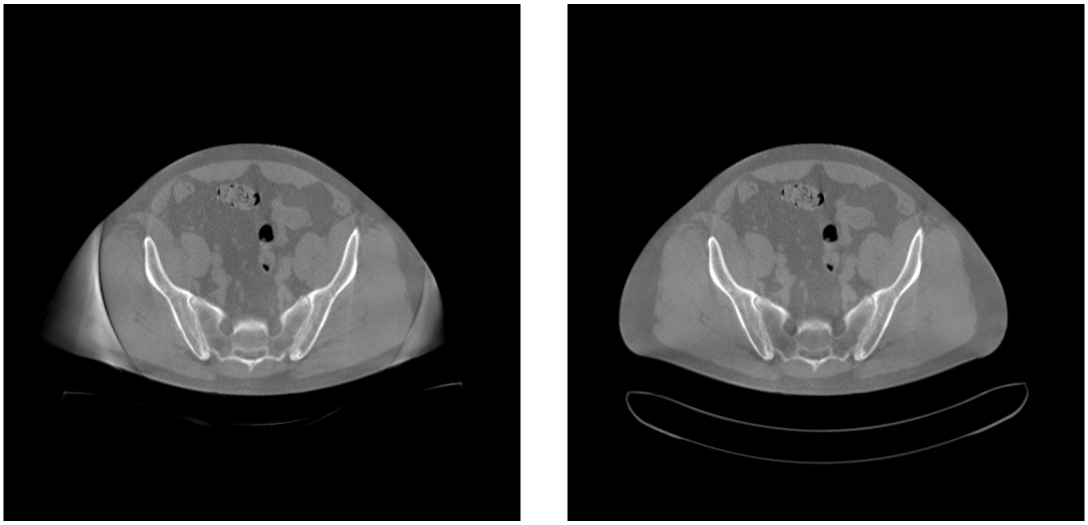

(a) (b)

Figure 4.5: One slice of abdominal scan reconstruction. Image window set to [0.005, 0.035] mm^{-1}. (a) Method 2 after 39 iterations; (b) Complete data reconstruction after 39 iterations.

truncated 125 detector elements (in each detector row) from both sides of the sinogram. This led to undersampling of the bed and a lateral portion of the patient's body. The left image was reconstructed using Method 2 on the truncated data. The qualitative appearance of the reconstructed image was similar to that which was observed in the unregularized simulated data experiment. For comparison, the image on the right was generated using the complete sinogram. A total of 39 full iterations were run using 145 ordered subsets to generate Figure 4.5.

4.3.2 Longitudinal Truncation

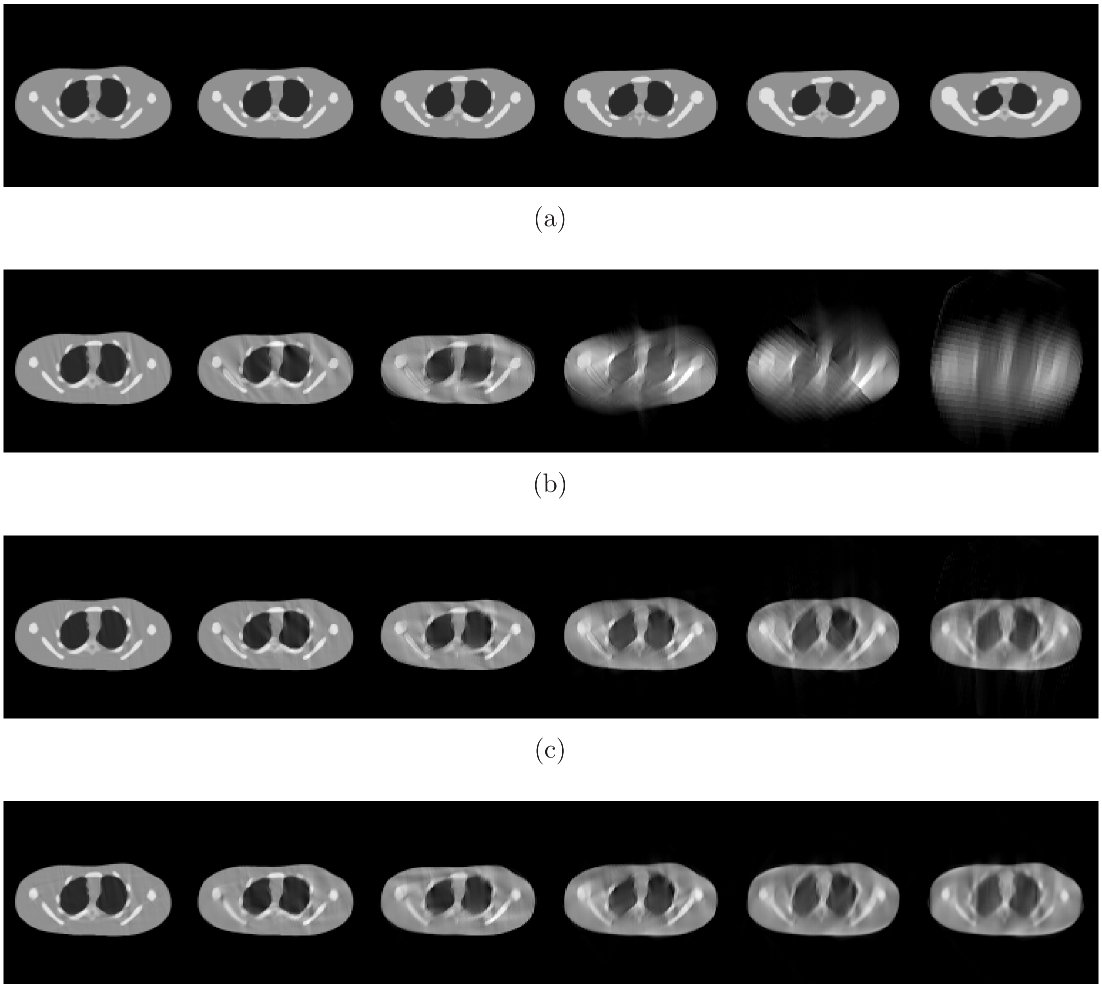
To study longitudinal truncation, we again use the NCAT phantom sampled with 168 detector elements per row, 8 detector rows, 584 views per rotation, and a pitch of 2. The true end slices of the reconstruction volume used to study the long object problem are shown in Figure 4.6(a). To demonstrate the appearance of artifacts due to the long object problem in Figure 4.6(b), we performed Method 1 without initializing the end slices as described in Section 4.2. Without prior information, the algorithm cannot perform a reasonable reconstruction because the end slices are so severely undersampled. When Methods 1 and 2 are initialized according to Section 4.2, we are able to improve the quality of slices closer to the end of the reconstruction cylinder, as seen in Figures 4.6(c) and 4.6(d), respectively.

However, careful inspection of the last one or two slices reveals that they are not very similar to the corresponding truth slices, but rather, are more similar to their initialized values (taken from slice 75). All reconstructed images in this figure were computed by running 100 full iterations. To account for the fact that fewer valid rays are available in each subset of $\mathcal{D}_{\text{miss}}$ than in each subset of $\mathcal{D}_{\text{inc}}$, we increased the number of source angles that are included in each subset of $\mathcal{D}_{\text{miss}}$. Specifically, we chose to use 73 source angles per rotation for each subset of $\mathcal{D}_{\text{miss}}$ while using eight source angles per rotation for each subset of $\mathcal{D}_{\text{inc}}$.

4.4 Discussion

The NCAT phantom and abdominal reconstructions using transversely truncated datasets demonstrate that it is possible to reconstruct the image inside the FOV quite accurately without many iterations. Outside the FOV, we again did reasonably well, although some artifacts are present. These artifacts diminish with increased numbers of iterations, but at a computational cost. We verified this by performing a 2D reconstruction on slice 70 of the NCAT phantom, in which case it was feasible to run thousands of iterations. Since this was a feasibility study, we omitted quantitative figures of merit to compare truth images to reconstructed images; once general improvements have been made to these methods, more in-depth studies can be performed.

If information in the extended FOV is needed for a particular application, then convergence rate appears to be the largest challenge in putting these methods into practice. If the goal is to perform ROI imaging, the reconstructions inside the FOV are consistently good after just tens of iterations. Regardless, to make these methods more clinically feasible, it will be necessary to look further into convergence acceleration techniques; as it is currently, each iteration of fully 3D reconstruction is still quite computationally demanding. Perhaps an additional prior penalty will help, or perhaps some more fundamental change to the algorithm can be made.



(a)

(b)

(c)

(d)

Figure 4.6: Truth image and unregularized image reconstruction results shown for end slices 76-81 using various methods. Image window set to [0.0, 0.07] mm^{-1}. (a) Truth image; (b) Method 1 without end slice initialization; (c) Method 1 with end slice initialization; (d) Method 2 with end slice initialization.

Based on Figure 4.2, Method 2 may have slightly better convergence properties than Method 1 for transverse truncation problems; this was also experimentally found in Snyder et al. [89]. However, Method 1 has the advantage that it is easier to implement and there are fewer computations compared to Method 2. One potential disadvantage of both Methods 1 and 2 is that the image space must be large enough to encompass all measured rays. For example, despite the larger required image space in our obese patient reconstruction, we decided to continue using 512×512 slices for the sake of computational convenience; this resulted in 1.4 mm voxels in the transverse plane rather than the original 1.0 mm voxels. The low contrast in Figure 4.4(a) could possibly be explained by this fact as well.

Viewing our proposed longitudinal truncation method from a computational standpoint, it would be preferable to estimate the missing projections using the same system matrix lookup table that is used in the computations over $i \in \mathcal{D}_{\text{inc}}$. However, since we are estimating the projection data on a computer (i.e., not physically), we are free to use a different source trajectory and/or finer detector sampling. For example, this would allow us to use a virtual circle trajectory at the ends of the reconstruction volume instead of our current helical trajectory (if desired). Although not tested here, there may be better sampling trajectories that can be employed for the purpose of improving Method 2.

Chapter 5

Reconstruction of Simulated Half-Ring PET Insert Datasets

The main purpose of the work described in this chapter was to develop a preliminary 2D reconstruction framework for the half-ring insert system while the physical half-ring prototype system was being built by our group. We focused on modeling the geometrical and attenuation factors of the system matrix, with the intention that these factors could be extended to 3D for use in the reconstructions described in Chapter 6. This work is itself an extension of the initial half-ring reconstruction approach described in Pal et al. [69], in that it now accounts for intercrystal penetration, crystal interference, and body attenuation; all three effects are important components in realistic models for PET data. It also models the native geometry of the insert system, unlike our previous work, which did not account for the detector block structure.

The organization of this chapter is as follows. Section 5.1 describes the 2D system geometry, the underlying data model, and the reconstruction algorithm that jointly estimates a single image from the three types of coincidence acquired by our system. Section 5.2 expands on each of the system matrix factors, and provides a derivation of the geometrical factor. We then briefly describe the MC simulations in Section 5.3. Finally, the experiments, results, and discussion are the subject of Sections 5.4-5.6, respectively.

Some material in Chapters 5 and 6 is based on "System Modeling of a DOI-Capable PET Insert Device for Breast Imaging," by D. B. Keesing, D. Pal, J. A. O'Sullivan, S. Komarov, and Y.-C. Tai, which appeared in the 2008 IEEE Nuclear Science Symposium Conference Record, 4218-4222;

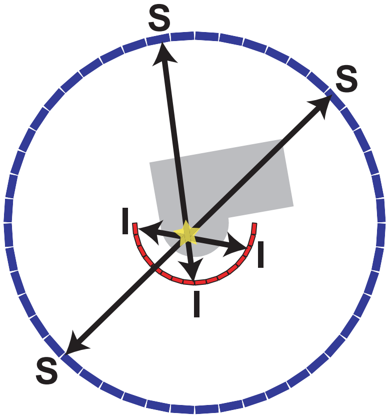


Figure 5.1: PET insert system geometry showing the II, IS, and SS coincidence types. Drawn to scale. Shown in gray is a body phantom, where one breast is positioned in the insert FOV.

5.1 Theory

5.1.1 System Geometry

Figure 5.1 shows the arrangement of our system, which consists of a half-ring of high-resolution detector crystals inside the FOV of a clinical scanner with medium-resolution crystals. Also shown in this figure is a body phantom with one breast positioned in the insert FOV. As indicated, three sets of coincidence data are acquired during a single scan: insert-insert (II), insert-scanner (IS), and scanner-scanner (SS). The II and IS datasets provide incompletely sampled data.

For the purpose of the MC simulations described in this chapter, the insert is made of 16 blocks each containing 16 crystals, and each crystal has dimension 1.6×1.6×10 mm. The blocks span 180° with radius 130 mm, resulting in a total of 256 crystals. The scanner is built using blocks of 13 crystals, with each crystal having dimension 4.15×1.6×20 mm. The blocks spanned 360° with radius 412 mm, resulting in a total of 624 crystals. The insert and scanner crystals were made of lutetium oxyorthosilicate (LSO) scintillation material.

Unlike conventional PET scanners, where it is acceptable to use parallel-beam rebinning of data, we cannot rebin data from our system since the IS geometry inherently acquires fan-beam projections between multi-resolution detectors. Given the complexity of our system, it is more natural to model the native geometry; this may lead to resolution improvements, and it eliminates the need to perform arc correction [12] to compensate for the non-uniformly spaced parallel beams.

5.1.2 Statistical Data Model

A pair of gamma ray photons are produced when a positron emitted from the radiotracer annihilates with a nearby electron. Given the nature of radioactive decay, the basic statistical model for the measured data assumes the photons arrive at the detector crystals according to a Poisson counting process. Let $d_{t,i}$ be the number of counts detected by a detector pair, where index $t \in \{0, 1, 2\}$ refers to coincidence type II, IS, and SS, respectively, and index i denotes a particular LOR within that coincidence type. In discretized form, the mean value of $d_{t,i}$ is modeled as

$$g_{t,i}(\boldsymbol{\lambda}) \triangleq E[d_{t,i}] \triangleq \sum_j p_{t,ij} \lambda_j, \tag{5.1}$$

where for coincidence type t, $p_{t,ij}$ is the probability that a gamma ray pair emitted from voxel j is detected in LOR i, and λ_j is the activity in voxel j. The summation in (5.1) is the forward projection of the image representing the true activity distribution. We did not consider scatter or randoms in the 2D MC simulations, which is why these additive terms do not appear in (5.1) as they had in (2.15).

Following the general strategy of Mumcuoğlu et al. [61] and Qi et al. [76], we factor the system matrix into a number of components. The factorization enables different effects to be calculated (or measured) separately, thereby preserving the maximum degree of sparseness in the system matrix. This will be especially important in the next chapter when we address 3D PET scanners, which have a very large number of possible LORs. Specifically, we factor the system matrix P as

$$P \triangleq P_{\text{norm}} \cdot P_{\text{atten}} \cdot P_{\text{geom}}, \tag{5.2}$$

where $[P]_{ij} \triangleq p_{ij}$. The t index was dropped for convenience, but note that there is a separate system matrix for each coincidence type. Alternatively, one can think of (5.2) as a generalized notation, in which the matrices corresponding to each coincidence type have been concatenated together to form one larger matrix; this is possible since all the data measurements are assumed to be independent and will be used in the same general manner in the reconstruction.

5.1.3 Image Reconstruction Formulation

The basic goal of image reconstruction in emission tomography is to estimate $\boldsymbol{\lambda}$ from $\boldsymbol{d}$. Most analytical reconstruction methods and rebinning methods cannot readily be extended to work with non-circular scanner geometries. Therefore, we take advantage of the flexibility offered by statistical reconstruction algorithms to handle the complexity of our system. More specifically, we seek to estimate a single emission image using all acquired data from the insert and scanner system.

The II, IS, and SS datasets are considered independent Poisson measurements, so at a high level, we have chosen to use the classic expectation-maximization (EM) algorithm [52, 87] to maximize the Poisson log-likelihood function with respect to the emission image voxels. The log-likelihood function for our system is given by

$$L(\boldsymbol{d}|\boldsymbol{\lambda}) = \sum_t \sum_i d_{t,i} \ln g_{t,i}(\boldsymbol{\lambda}) - g_{t,i}(\boldsymbol{\lambda}) - \ln d_{t,i}!, \tag{5.3}$$

and the maximum likelihood image estimate is therefore

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\lambda \geq 0} \sum_t \sum_i d_{t,i} \ln g_{t,i}(\boldsymbol{\lambda}) - g_{t,i}(\boldsymbol{\lambda}). \tag{5.4}$$

The EM algorithm (adapted to the three types of coincidence) maximizes the expression in (5.4), and is given by

$$\hat{\lambda}_j^{(k+1)} = \frac{\hat{\lambda}_j^{(k)}}{\sum_t \sum_i p_{t,ij}} \sum_t \sum_i p_{t,ij} \frac{d_{t,i}}{\sum_j p_{t,ij} \hat{\lambda}_j^{(k)}}. \tag{5.5}$$

As it is written, the denominator $\sum_t \sum_i p_{t,ij}$ serves as a normalization factor for each voxel. The main part of the image update step is the backprojection of the ratio between the uncorrected simulated true counts and the mean estimate of the true counts at iteration k.

5.2 System Model

One challenge for our system geometry is the fact that both the front face and the lateral faces of the insert crystals are directly exposed to gamma rays emitted from the patient. This is in contrast to most cylindrical PET scanners, in which the geometrical component of the system matrix can be accurately modeled by determining the probability that an emission in a given voxel reaches just the front faces of the crystal pair [76].

The method presented in [69] for calculating the system matrices was to first subdivide the crystal faces into small areas. For the IS type, the insert crystals were subdivided along the front and lateral faces, while the scanner crystals were subdivided just along the front face. Next, sub-rays were formed by connecting the center of one subface on the first crystal to the center of a subface on the second crystal. Finally, the weight for a given pixel/LOR pair was determined by taking the average length of intersection of the pixel with all combinations of the sub-rays.

Although this approach accounted for activity above the side face of the insert crystals, it was not a very realistic model, in that the sensitivity calculated at each face was weighted improperly before adding the sensitivities from multiple faces together. This was particularly true when the data was attenuated by the body (which is always the case in reality except in some initial MC simulations, where the body was modeled as air.)

5.2.1 Geometrical Factor

The geometrical effects related to the incidence angle of the LOR with respect to the front face of the detectors are often accounted for in the normalization factor [58]. This works well for circular PET scanners, but the methodology used to compute these effects cannot readily be extended to the II or IS coincidence types. We will therefore account for the LOR incidence angle effects within the calculation of P_{geom}, along with the image to data space transformation.

In Figure 5.2, we show a large plane source oriented perpendicular to two detectors, d_1 and d_2. The activity at point (x, y) within the plane source is seen by the LOR with coincident solid angle approximately proportional to the small shaded area. We define the coincidence response function of the LOR (based on solid angle considerations) as

$$\xi_\Omega(l) \triangleq \int\int \Omega(x, y, l)\, dx\, dy, \tag{5.6}$$

where $\Omega(x, y, l)$ is the coincident solid angle subtended by the point (x, y) within the plane identified by length l, and the integral is performed over the surface of the plane source. If the size of the detectors is much smaller than the length of the LOR, then it can be shown [37, 55] that $\xi_\Omega(l)$ is approximately constant regardless of the plane source location along the LOR, i.e., $\xi_\Omega(l) \approx \xi_\Omega$.

This approximation leads to the derivation of P_{geom}. Based on solid angle principles,

$$\xi_\Omega \approx \frac{A_1 A_2}{2\pi L^2}, \tag{5.7}$$

where A_i is the projected area of crystal $i \in \{1, 2\}$ onto the plane perpendicular to the LOR, and L is the distance between the crystals.

For a continuous object attenuation distribution $\mu(l)$ and emission distribution $\lambda(l)$ along the LOR, Huesman et al. [37] showed that as long as these distributions do not vary much along the perpendicular direction to the LOR, the coincidence response is

$$\xi_\Omega \approx \frac{A_1 A_2}{2\pi L^2}\, e^{-\int \mu(l)\, dl} \int \lambda(l)\, dl. \tag{5.8}$$

Huesman et al. [37] then accounted for the depth of interaction (DOI) within the crystal volumes by weighting the coincidence response, which up until this point, has been based solely on solid angle considerations. Letting ξ be the new coincidence response,

$$\xi \approx \xi_\Omega \int_{L_1}^{L_1'} \mu_{\text{det}} e^{-\mu_{\text{det}} t_1}\, dt_1 \int_{L_2}^{L_2'} \mu_{\text{det}} e^{-\mu_{\text{det}} t_2}\, dt_2\,, \tag{5.9}$$

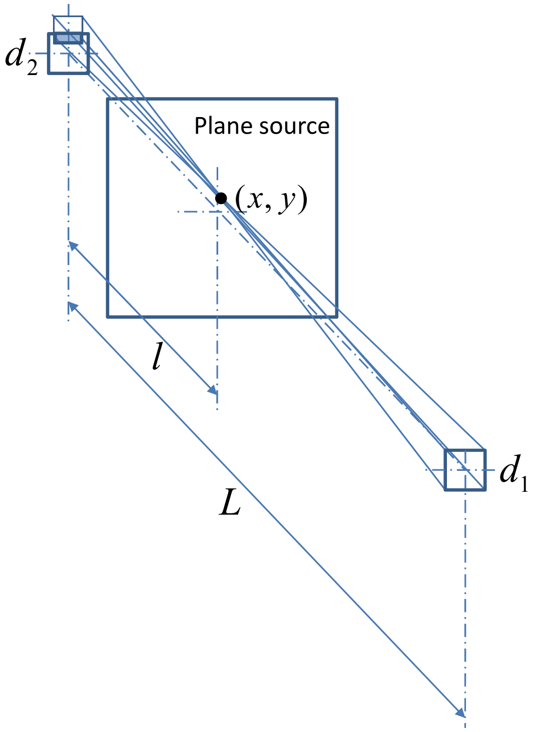


Figure 5.2: Plane source oriented perpendicular to an LOR formed by detectors d_1 and d_2. The small shaded region corresponds to the solid angle that is common to the detector pair for the point (x, y). In this diagram, both detectors are intended to be the same size and $l < L/2$, thereby making d_1 the solid angle limiting detector. (Based on figure by Dr. Sergey Komarov.)

where L_i is the length the gamma ray travels through detector material in front of the crystal of interest, L_i' is the total length the gamma ray travels through detector material including the crystal of interest, and μ_{det} is the linear attenuation coefficient of the detector material. This is identical to the weighting used in Yamaya et al. [108] (previously discussed in Chapter 2), although here it is being used in a slightly different context.

Once the emission distribution has been discretized into voxels, the final coincidence response becomes [37]

$$\begin{aligned}\xi &\approx \frac{A_1 A_2}{2\pi L^2}\, e^{-\int \mu(l)\, dl} \int_{L_1}^{L_1'} \mu_{\text{det}} e^{-\mu_{\text{det}} t_1}\, dt_1 \int_{L_2}^{L_2'} \mu_{\text{det}} e^{-\mu_{\text{det}} t_2}\, dt_2 \sum_{j=1}^{N} l_j \lambda_j \\ &= \frac{A_1 A_2}{2\pi L^2}\, e^{-\int \mu(l)\, dl} e^{-\mu_{\text{det}}(L_1+L_2)} \left[1 - e^{-\mu_{\text{det}}(L_1'-L_1)}\right] \left[1 - e^{-\mu_{\text{det}}(L_2'-L_2)}\right] \sum_{j=1}^{N} l_j \lambda_j \\ &\approx \frac{A_1 A_2}{2\pi L^2}\, e^{-\int \mu(l)\, dl} e^{-\mu_{\text{det}}(L_1+L_2)} \mu_{\text{det}}^2 (L_1' - L_1)(L_2' - L_2) \sum_{j=1}^{N} l_j \lambda_j,\end{aligned} \tag{5.10}$$

where l_j is the length of intersection of the LOR with voxel j, λ_j denotes the constant activity inside voxel j, and N is the number of voxels in the reconstruction FOV. The second approximation comes from the first-order Taylor series for the exponential function, which holds when the detectors are very small.

The detectors modeled above were also assumed to be very small for the spatially-invariant plane source coincidence response approximation to hold. As shown in Figure 5.3, we therefore subdivide the full crystal volume into subvolumes, with sub-rays joining each combination of subvolumes [37]. The coincidence response for the complete detector pair is then the sum of the coincidence responses for all sub-rays. For layered crystals (as in the case of DOI-capable detectors), only those subvolumes that constitute a given layer are summed. It is apparent that the crystal subvolume approach is more natural for DOI crystals than our previous crystal subface approach.

In the context of our reconstruction framework, the body attenuation is treated as a separate factor. Therefore, to fit the expression for ξ into our geometric forward projection operator, P_{geom}, we write

$$[P_{\text{geom}}]_{ij} \triangleq \frac{e^{-\mu_{\text{det}}(L_{i1}+L_{i2})}}{{L_i}^2}\, l_{ij}, \tag{5.11}$$

where l_{ij} is the length of intersection of LOR i with voxel j, and the expression is implicitly summed over all sub-rays comprising the LOR. As in Huesman et al. [37]

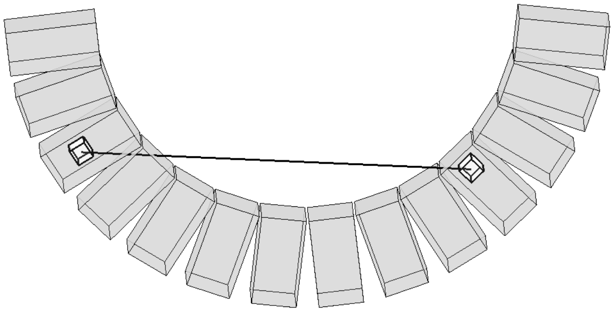


Figure 5.3: Subdivision of crystals to calculate the geometric system matrix elements. Note that the ray passes through other crystal material (and air between blocks) before reaching the two crystal subvolumes of interest.

and Hu et al. [35], the expression

$$\frac{\mu_{\text{det}}^2}{2\pi} A_1 A_2 (L_1' - L_1)(L_2' - L_2) \tag{5.12}$$

is assumed to be constant for small enough subdetectors. In our method, this approximate constant is accounted for through the separate P_{norm} factor.

5.2.2 Body Attenuation Factor

In our current implementation, the gamma ray survival probability (after attenuation by the body) is calculated along each LOR as

$$[P_{\text{atten}}]_{ii} \triangleq \exp\left(-\sum_{j=1}^{N} a_{ij}\mu_j\right), \tag{5.13}$$

where a_{ij} is the geometric system matrix for transmission forward projections, and μ_j is the linear attenuation coefficient (at 511 keV) at voxel j of the body attenuation map. In this work, a_{ij} was computed by taking the average length of intersection of voxel j with LOR i, where the average was taken over all subvolume combinations within the crystal pair. To determine P_{atten} for the MC simulations performed in this

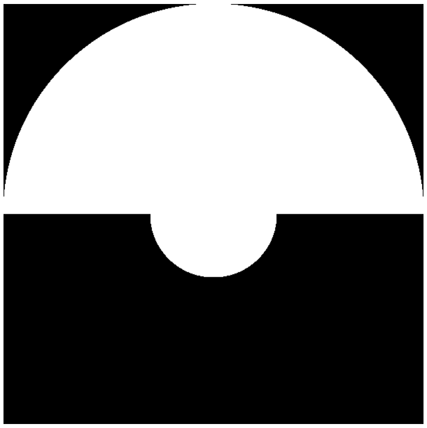


Figure 5.4: Image of the phantom used for direct normalization.

work, the analytical specifications for the body were voxelized into an attenuation map with the same dimensions as the emission image.

5.2.3 Normalization Factor

We have developed a normalization procedure to account for the insert attenuation and any other effects that are not modeled by the system matrix factors. For example, although the crystals had the same intrinsic efficiency in our simulations, variations in efficiency could be accounted for using this normalization factor.

The diagonal normalization factor is calculated using a direct normalization approach, in which a known source distribution is used to illuminate all LORs in the FOV. In this work, we use the uniform source distribution shown in Figure 5.4, which is large enough to cover all LORs that would ever be illuminated by arbitrary body source distributions. The normalization activity is suspended in air in the simulation so attenuation does not need to be calculated.

To obtain the normalization factor, we first run a high-statistics MC simulation using the uniform normalization phantom. Next, we calculate the forward projection through the voxelized normalization phantom image using the P matrix with all factors except for P_{norm}. The normalization factor for each detector pair is then determined by taking the ratio of the MC simulation data to the forward projection.

5.3 Monte Carlo GATE Simulations

We use the GATE software package (Geant4 Application for Emission Tomography) [56] to perform the MC simulations of our insert PET device. This is well-validated code for simulating medical imaging scanners, and is built on top of Geant4, a lower-level MC engine developed at CERN. GATE gives the user a large degree of control over the simulation parameters through scripts and macros that the user creates. Based on these scripts, GATE can simulate arbitrary systems with as many or few physical effects as desired, such as the detector electronics processing chain.

For our insert system, we have to perform each simulation twice, due to the fact that only one set of active detectors is allowed at a time. Therefore, when simulating data recorded by the insert detectors, the scanner is considered a phantom ring of LSO crystals; when simulating data recorded by the scanner detectors, the insert is considered a phantom half-ring of LSO crystals. The random number generator seed is kept the same so the events generated in both simulations are identical.

The 2D simulations were set up such that all the objects and detectors had a thickness equal to the thickness of a single insert crystal in the axial direction of the scanner (1.6 mm). Sorting was performed by annihilation event ID rather than by time-based coincidence processing, so there were no randoms included in the dataset. Additionally, we used a 510 keV lower energy threshold and assumed the detectors had perfect energy resolution, so there were no scattered counts included in the dataset either.

In addition to the elimination of randoms and scattered counts, many other effects were omitted from these simulations as well, in order to focus on the most important aspects of the 2D system model and the reconstruction procedure itself. For example, there were no alignment errors between the normalization phantom image and the actual phantom used in the simulation. Similarly, the insert was centered exactly in the scanner. The simulations were accelerated by simulating back-to-back gamma rays instead of positron emissions (i.e., positron range and gamma ray acolinearity were not simulated), and the gamma rays were forced to be emitted in-plane for efficiency.

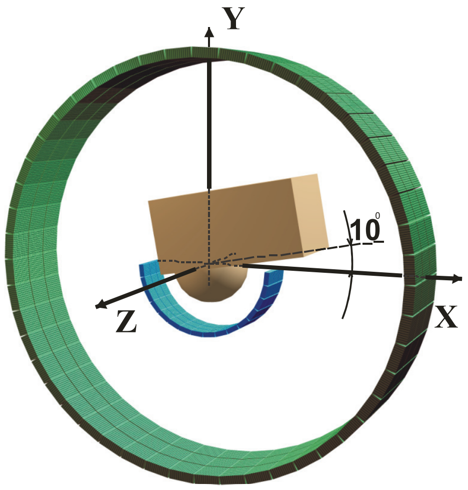


Figure 5.5: Breast phantom. (Figure courtesy of Dr. Sergey Komarov.)

5.4 Experiments

To compare the performance of the subvolume approach to that of the subface approach for computing P_{geom}, we simulated one slice of the body phantom shown in Figure 5.5. Insert and scanner crystals were subdivided four times in the tangential direction, and eight times in the depth direction. The chest was modeled as a 30×15×0.16 cm box rotated 10° from horizontal. The breast was modeled as a half-cylinder with radius 7.5 cm and thickness 0.16 cm. Activity was distributed uniformly throughout the phantom, and the phantom material was specified to be water. Once the simulation was run, the data was sorted into II, IS, and SS sinograms, and subsequently reconstructed using the EM algorithm. One reconstruction was performed with the subface-based system matrix, and the other, with the subvolume-based system matrix. The EM iterations were initialized with the normalization phantom image; as can be seen in (5.5), the initial image must be positive to ensure the image update does not simply get multiplied by a zero image.

To further test the subvolume approach, we ran a simulation of eleven 0.5 mm diameter point sources placed horizontally and vertically in the insert FOV. The data was first sorted into II, IS, and SS sinograms as done above. However, we also sorted the

---

Although the 3D model of the breast in Figure 5.5 is hemispherical, the simulated slice was sufficiently thin that it was assumed to have a cylindrical cross section.

data into sinograms corresponding to a four-layer DOI insert system, where each layer in the insert crystals was 2.5 mm thick. This increased the number of II sinograms ten-fold, and the number of IS sinograms four-fold. Sorting of the events into the appropriate DOI sinograms was done based on the spatial coordinates where each gamma ray interacted with the detector volume (as recorded in the detector singles file). The main reason for testing DOI insert detectors was to determine how well it could reduce the parallax error that results when sources are far from the center of the FOV. Since the body can come much closer to the edge of the insert FOV than it can to the edge of the scanner FOV, it is expected that parallax error could be significant in the insert.

5.5 Results

5.5.1 Body Phantom Simulation

Figure 5.6(a) shows the reconstruction of the body phantom using the subface-based system matrix, and Figure 5.6(b) shows the reconstruction using the subvolume-based system matrix. As can be seen, significant artifacts are produced when using the subface approach with body-attenuated data. These artifacts are avoided using the subvolume method.

The data sinograms from the body phantom simulation are displayed in the left column of Figure 5.7 for II, IS, and SS. For comparison, the mean forward projections from the last iteration of the subvolume-based reconstruction are displayed in the right column of Figure 5.7.

5.5.2 Point Source Simulation

Figure 5.8 shows the reconstructions of the point source data using non-DOI insert crystals and four-layer DOI insert crystals. These reconstructions were done only

---

The increase in the number of sinograms corresponds to the number of unique layer combinations that exist between crystal types.

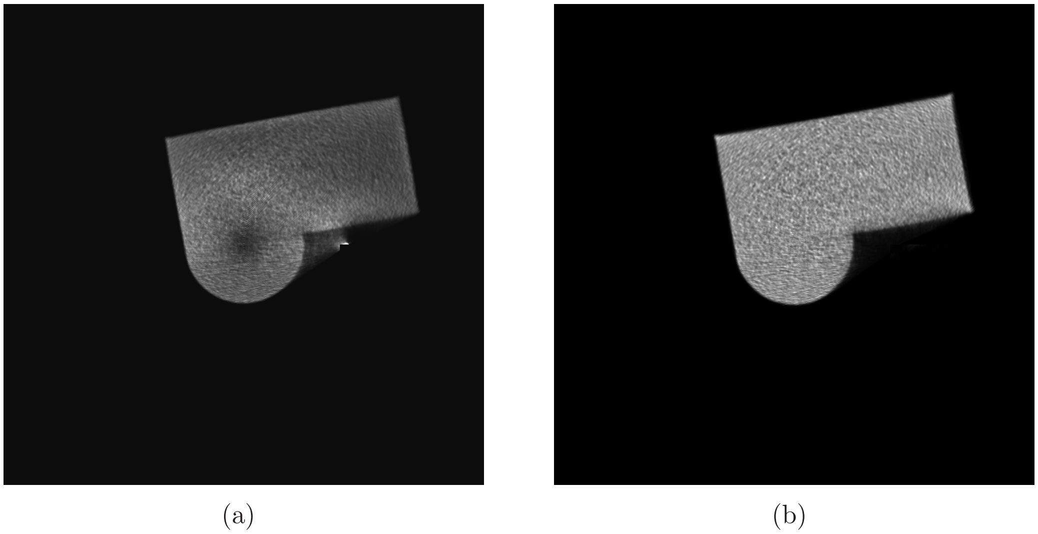

(a) (b)

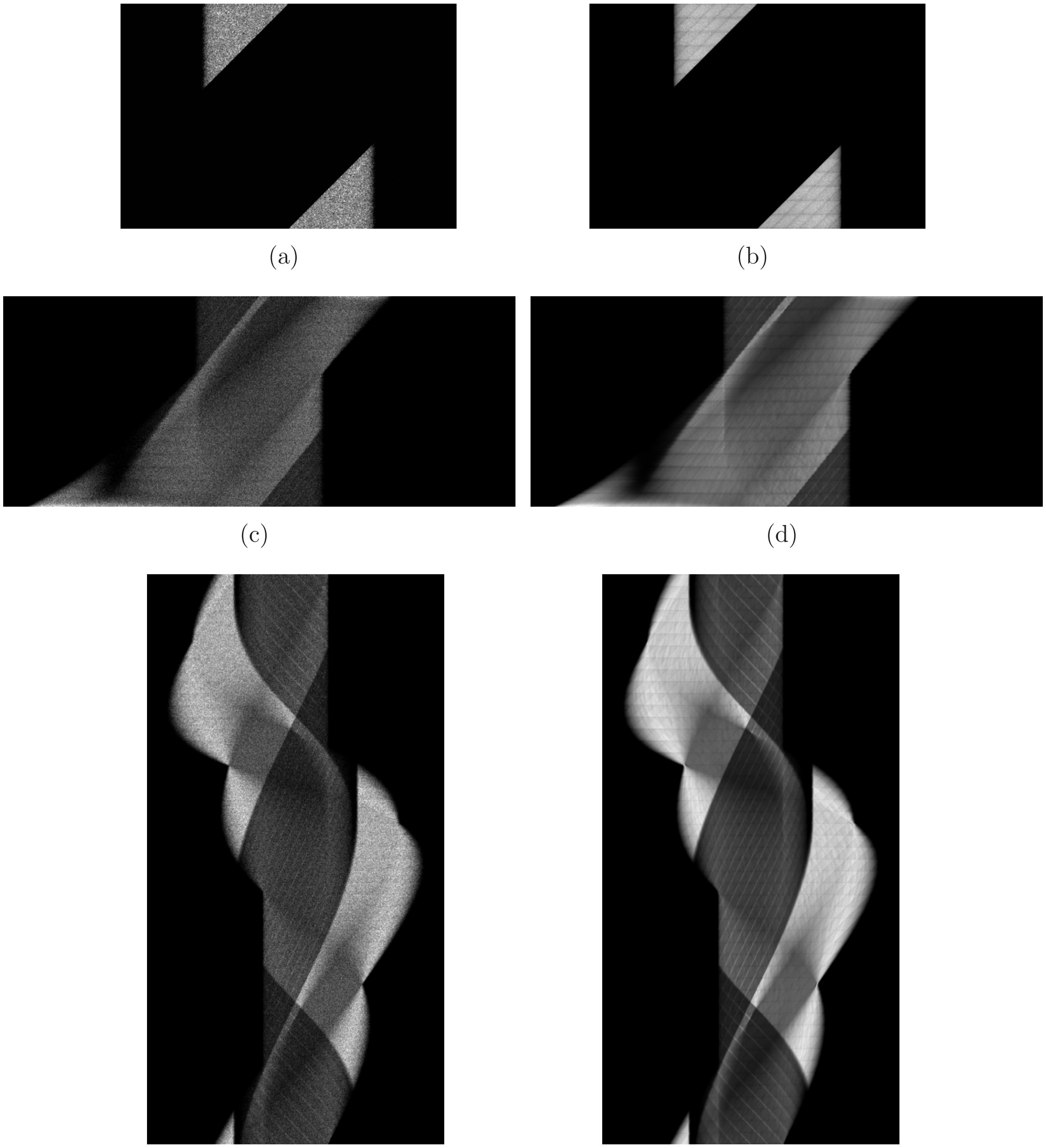
Figure 5.6: II+IS+SS reconstructions of the body phantom after 100 EM iterations. (a) Reconstruction in which P_{geom} was calculated using the crystal subface approach; (b) reconstruction in which P_{geom} was calculated using the crystal subvolume approach.

using the subvolume approach. The images demonstrate that parallax error is an issue for sources near the bottom edge of the insert FOV in the non-DOI insert system, as seen by the elongated point sources. As expected, the DOI insert system can effectively remove such radial distortions.

In Figure 5.9, we plot profiles through the point sources that were arranged vertically; the left side of the profile corresponds to the upper point sources. It is clear from these profiles that the point source resolution is higher for the DOI reconstruction.

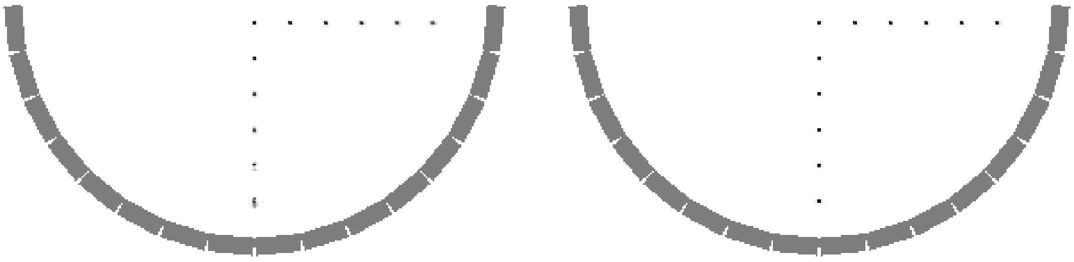
5.6 Discussion

In this chapter, we have established the data model and framework for reconstructing 2D simulated half-ring insert PET data. We have applied the main aspects from Huesman et al. [37] to the geometrical factor of our system matrix. It includes weights that account for the intercrystal penetration and crystal interference. By systematically accounting for these effects in the geometrical factor, we avoid needing to compute them in the normalization factor using a different type of method that



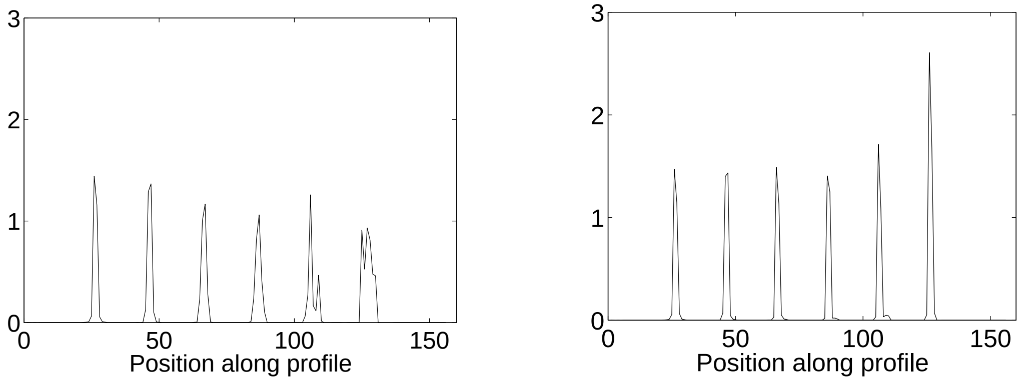
(a) (b) (c) (d) (e) (f)

Figure 5.7: Comparison of sinograms for the simulated body phantom. Top row: II, middle row: IS, bottom row: SS. Left column: MC simulated data, right column: mean forward projection after 100 EM iterations.



(a) (b)

Figure 5.8: II+IS+SS reconstructions of point sources using (a) non-DOI insert crystals; (b) 4-layer DOI insert crystals. The insert blocks have been overlaid on the reconstructed images for reference.



(a) (b)

Figure 5.9: Profiles through reconstructed images of point sources in the II FOV. The left side of the profile corresponds to the upper portion of the FOV. (a) non-DOI insert crystals used for data acquisition and reconstruction; (b) four-layer DOI insert crystals used for data acquisition and reconstruction.

does not readily apply to the half-ring insert. This geometrical method, which is based on subdividing the crystals into subvolumes, can also naturally handle DOI detectors.

The simulations that we performed in this chapter modeled the insert as consisting entirely of LSO crystals, e.g., the simulations did not include photomultiplier tubes or an insert cover. We were able to employ a simple direct normalization method since we did not record randoms or scattered counts.

It was shown that the subface method for computing P_{geom} leads to artifacts in the reconstructed images. In previous work, we had incorporated the body attenuation factor directly into the geometric factor calculations over all sub-rays to avoid such artifacts. We could do this very accurately since the shape and attenuation parameters for the body were specified in the MC simulations. Although this method of modeling the system and body worked reasonably well, it was not practical, as it removed all symmetry from the geometrical factor. Instead, by using the subvolume-based approach derived in this chapter for the geometric factor, along with a separate diagonal body attenuation factor, it is possible to avoid artifacts in the reconstructions.

We compared the measured and mean data sinograms, which provides direct insight into how accurately the data is modeled and normalized. The fact that these sinograms look so similar suggests that the system matrix (consisting of the geometric, attenuation, and normalization factors) is indeed able to sufficiently model the simulated data.

As discussed in the derivation of P_{geom}, the subvolume approach assumes there is not significant variation in the attenuation perpendicular to a given LOR. In the case of LORs pairing the side face of an insert crystal to the scanner, this assumption might not be entirely valid since the side face has a much broader area than the front face. The use of DOI insert crystals would help alleviate this problem since they narrow the solid angle over which counts for a given LOR are detected. The point source results shown here simply indicate that DOI crystals are able to reduce the parallax error that occurs at LORs far from the center (due to the increased angle of incidence the LOR makes with the crystals).

Chapter 6

Reconstruction of Half-Ring PET Insert Datasets Acquired with Physical Prototype

In this chapter, we present the details of the PET half-ring insert prototype and the reconstruction procedure that has been developed. As previously discussed in Chapter 5, a new type of PET insert system has been developed and integrated into a whole-body PET scanner. It offers the possibility of imaging a central region of the FOV with higher resolution than a whole-body PET scanner alone [95, 96]. Unlike most dedicated high resolution PET systems that are designed solely to image one part of the body, our system maintains the FOV that includes the entire cross-section of the body, thereby potentially allowing for detection of tumors inside and outside the high-resolution insert FOV.

The data acquired by the physical system includes many sources of degradation, mainly due to the presence of the highly attenuating insert system within the imaging FOV of the scanner. This novel scanner geometry makes the reconstruction problem substantially more challenging than for a standalone PET scanner, as new techniques are needed to model and correct for the non-standard acquisition. Compared to the full-ring insert system [68], the half-ring system presents additional complications due to its incomplete sampling geometry and lack of circular symmetry.

In the MC simulations discussed in Chapter 5, we were able to simplify the reconstruction problem to some extent by disabling or removing some of the effects that degrade the data. There were no physical alignment issues to consider either, since the detector geometry was specified exactly. It is difficult if not impossible to prevent these effects from negatively affecting the physically measured data, however.

The organization of this chapter is as follows. We will explain the geometry of the system, along with a description of the chosen data model and reconstruction method, in Section 6.1. Section 6.2 expands on each of the system matrix factors, and establishes the methodology by which randoms and scatter are estimated. The major emphasis in this section is the normalization factor. Section 6.3 highlights the computational aspects of the reconstruction procedure. The prototype system and procedure for calculating the attenuation due to the insert is described in Section 6.4. We present a number of phantom scans and their results in Sections 6.5 and 6.6. Finally, a discussion is provided in Section 6.7.

6.1 Theory

6.1.1 System Geometry

Figure 6.1 shows the arrangement of our system, which consists of a semicircular ring of high-resolution detectors inside the FOV of a clinical scanner with medium-resolution crystals. By integrating the insert detector electronics with the clinical scanner hardware, three sets of coincidence data are acquired simultaneously: insert-insert (II), insert-scanner (IS), and scanner-scanner (SS). We used a Biograph 40 PET/CT scanner (Siemens Medical Solutions, Knoxville, TN), which has 48 blocks of crystals in-plane and 4 blocks axially. Each block consists of 13×13 LSO crystals, and each crystal is $4 \times 4 \times 20$ mm. As shown in Figure 6.1, our system only uses three axial blocks of scanner crystals; the electronics for the last block are instead used by the insert detectors.

The half-ring insert gantry has 14 blocks of crystals in-plane and 2 blocks axially. Each block consists of 13×13 LSO crystals, and each crystal is $2 \times 2 \times 5$ mm. The radius to the center of the scanner blocks' front face is 428 mm, and the radius to the center of the insert blocks' front face is 122 mm. The insert is centered axially over the three active scanner ring blocks, and in the in-plane direction, it is concentrically

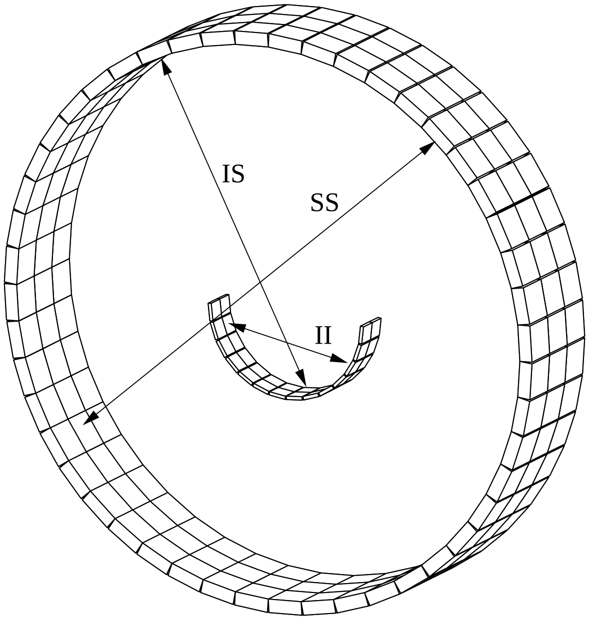


Figure 6.1: PET insert system geometry showing the II, IS, and SS coincidence types. Drawn to scale.

centered within the scanner FOV. Since the insert crystals have the same physical dimensions in the transverse and axial directions, the benefits of the insert VP-PET system apply to both directions.

6.1.2 Statistical Data Model

We use the same general statistical data model as in Chapter 5, except now we also must account for randoms and scattered counts. These background counts are unavoidable during acquisition, but can be estimated and incorporated into the data model. Again, let $d_{t,i}$ be the number of counts detected by a detector pair, where index $t \in \{0, 1, 2\}$ refers to coincidence type II, IS, and SS, respectively, and index i denotes a particular LOR within that coincidence type. In discretized form, the mean value of $d_{t,i}$ is now modeled as

$$g_{t,i}(\boldsymbol{\lambda}) \triangleq E[d_{t,i}] \triangleq \sum_j p_{t,ij}\lambda_j + \bar{r}_{t,i} + \bar{s}_{t,i}, \tag{6.1}$$

where for coincidence type t, $p_{t,ij}$ is the probability that a gamma ray pair emitted from voxel j is detected in LOR i, λ_j is the activity in voxel j, $\bar{r}_{t,i}$ is the mean number of random coincidences detected by the LOR, and $\bar{s}_{t,i}$ is the mean number of scattered coincidences detected by the LOR. The first term in (6.1) is the forward projection of the image representing the true activity distribution.

To maintain the validity of the Poisson model, we do not pre-correct the prompt data for randoms by subtracting the delayed counts. If the subtraction were performed, the measured data variance would increase by the delayed counts variance, while the mean would decrease by the delayed counts mean. The mismatched mean and variance would no longer be in accordance with the Poisson distribution, thereby necessitating the use of another distribution, such as the shifted Poisson model [110]. Instead, as shown in (6.1), adding a relatively smooth background estimate to the forward projection avoids this complication.

As in the previous chapter, the system matrix P is factored as

$$P \triangleq P_{\text{norm}} \cdot P_{\text{atten}} \cdot P_{\text{geom}}, \tag{6.2}$$

where $[P]_{ij} \triangleq p_{ij}$. The factors serve the same purpose as before, but will be determined differently for the physical system.

6.1.3 Image Reconstruction Formulation

Due to the general nature of statistical reconstruction algorithms, the same EM algorithm can be used regardless of whether the reconstruction is to be performed in 2D or 3D. The nonzero randoms and scattered counts contribution to the data model is reflected in the denominator of the EM image update given by

$$\hat{\lambda}_j^{(k+1)} = \frac{\hat{\lambda}_j^{(k)}}{\sum_t \sum_i p_{t,ij}} \sum_t \sum_i p_{t,ij} \frac{d_{t,i}}{\sum_j p_{t,ij} \hat{\lambda}_j^{(k)} + \bar{r}_{t,i} + \bar{s}_{t,i}}. \tag{6.3}$$

The main part of the image update step is the backprojection of the ratio between the uncorrected measured prompts data and the mean estimate of the prompts data at iteration k. Note that the prompts data includes trues, randoms, and scattered counts. We assume the mean randoms and scatter estimates are accurately known in advance. If this is not the case, the estimates can be refined between iterations using various update strategies.

6.2 System Model

Determination of the factors that constitute P for the real prototype system require more extensive work than for the MC simulations, and will be the focus of this section. A method to estimate the randoms and scattered counts will also be described.

6.2.1 Geometrical Factor

We have already derived the geometrical factor in Section 5.2.1. This factor is general enough to apply to the 3D system by calculating (5.11) for the oblique detector pairs. Due to the irregular sampling of the insert system, we compute the system matrix using the native detector geometry. To do this, the 3D position and orientation of every crystal and block is used in the calculations. This is unlike many other reconstruction algorithms that simplify the geometry by using parallel-beam projections taken around a smooth circle. As before, the number of subdivisions within each crystal can be adjusted to achieve the desired accuracy. Since the main computation in (5.11) is the voxel-ray intersection length calculation, the geometrical factor can be calculated efficiently using a fast voxel traversal algorithm [3]. Calculation of L_{i1} and L_{i2} was done using a fast oriented bounding box/ray intersection method [2] with the calculated detector block geometry.

6.2.2 Body Attenuation Factor

The P_{atten} factor only accounts for the attenuation due to the body in our current implementation; the insert attenuation is treated separately in P_{norm}, since it does not change from scan to scan. Since the insert is integrated with a clinical PET/CT scanner, an x-ray CT attenuation map of the body (scanned without the insert in

Parameter	Value
Tube voltage	120 kVp
Tube current	(depends on object)
Pitch	1.0
Convolution kernel	B19f
Number of voxels	512×512×111
Reconstruction diameter	500 mm
Slice thickness	2 mm

Table 6.1: CT scan and reconstruction protocol for the body attenuation correction factor.

the FOV) can be obtained. To determine P_{atten}, we first perform a CT scan of the object using a protocol whose main parameters are specified in Table 6.1. The CT data is reconstructed on the Siemens Biograph 40 console, and subsequently saved to standard DICOM image files as an affine transform of Hounsfield units.

Based on an assumed effective energy of about 80 keV, we convert the attenuation image to units of mm^{-1}. This conversion estimates the linear attenuation coefficients at an effective energy of 80 keV, which is valid for CT but not for PET. We determine the attenuation coefficients at the gamma ray energy of 511 keV using a piecewise linear transformation [45, 11]. This standard approach performs an appropriate conversion for soft tissue and bone. The image dimensions and voxel size are then rescaled to those of the PET image. Finally, the attenuation factor is computed in the same manner as in Chapter 5, i.e.,

$$[P_{\text{atten}}]_{ii} \triangleq \exp\left(-\sum_{j=1}^{N} a_{ij}\mu_j\right). \tag{6.4}$$

The a_{ij} factors are calculated without subdividing the depth direction of the crystals in our current implementation. Based on our 2D MC simulation work, the number of subdivisions might need to be increased if large phantoms that extend over the side face of the insert are scanned.

6.2.3 Normalization Factor

We have investigated two approaches to compute the normalization factor for the real system. The first one uses direct normalization; it is similar to the method described in Chapter 5, but uses a more practical source distribution. The second approach uses component-based normalization, which requires a model for the variations in sensitivity among all the LORs.

Direct Normalization

The direct normalization method is designed to be simple to use, as it does not require a model of the individual normalization effects. For the insert system, this would eliminate the need to explicitly determine the attenuation due to the insert gantry. A rotating line source or plane source is typically used to illuminate all LORs in the FOV during the direct normalization acquisition. Next, a digital image of the source distribution is created and forward-projected using the modeled system matrix factors (besides P_{norm}). Finally, the ratio of the measured data to the forward-projected data is used as the normalization factor.

This method suffers from several important drawbacks, however. First, in a fully 3D PET acquisition without axial compression, the statistics are not very good (even after a long scan), since the source activity is kept low to avoid dead time effects. Second, it is difficult to align the source exactly the same way between the real system and the digital image. Especially for a thin rotating rod source, a small error in positioning can lead to a large error in the normalization factors. This is less of an issue for scanners that have a built-in holder for rotating line sources. Lastly, the treatment of scattered counts is not included in the direct normalization approach. Although the normalization phantoms themselves are designed to minimize the scatter volume, the scattered gamma rays arising from the half-ring insert may complicate the issue.

The large uniform normalization phantom that was used in the 2D MC simulations is not practical for the real system for a number of reasons: (1) it would be difficult to

---

Direct normalization can also be used to precorrect the measured data (i.e., prior to reconstruction), in which case the *inverse* ratio is applied to an unnormalized dataset to artificially make all LORs equally sensitive [58].

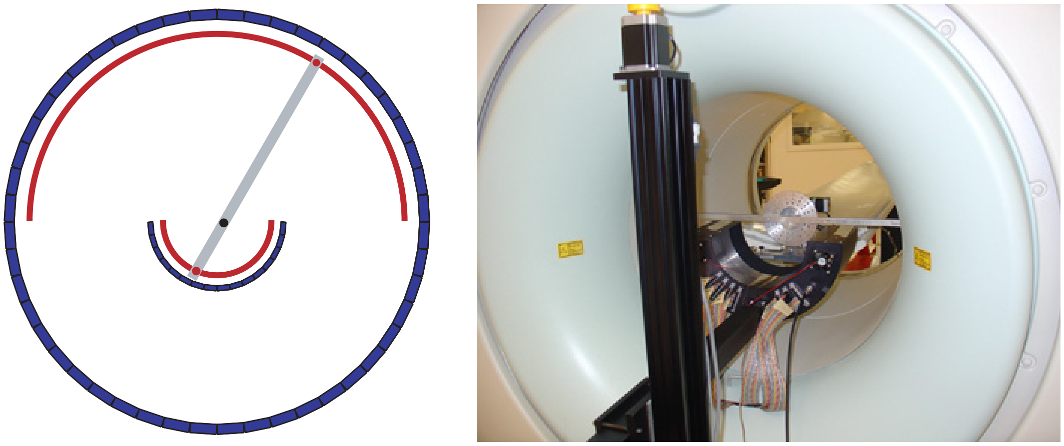

(a) (b)

Figure 6.2: The normalization phantom consists of two line sources (oriented out of the page) that are rotated 180° around the isocenter. (a) The rotation stage has its center of rotation at the black dot, and sweeps out the source distribution shown in red; (b) photograph of the rotation stage inside the insert and scanner FOV.

maintain a uniform activity distribution throughout the large phantom, (2) the scatter fraction would be quite large, and (3) it would be too expensive and heavy (when filled with water and ^{18}F-FDG). Instead, we used two line sources placed on a motor-controlled rotation stage, as shown in Figure 6.2. With the rotation stage centered at the system isocenter, we rotated the line sources back and forth 180° to sweep out the source distribution shown in red in the diagram. This source distribution is able to illuminate all LORs.

Component-Based Normalization

To avoid some of the problems with direct normalization, we also tested a component-based approach, in which each source of LOR sensitivity variation (other than those already accounted for in P_{geom} and P_{atten}) is modeled as a normalization component. Since there are far fewer parameters to estimate than the total number of LORs, it can effectively produce a normalization factor that has lower variance than direct normalization. However, bias may be inadvertently introduced if the normalization components are not modeled or computed appropriately.

Component-based normalization models of varying complexity have been proposed, e.g. [12, 6, 4]. Among the main components are single crystal efficiencies and geometrical effects, which account for intrinsic detector efficiency and sensitivity changes as a function of angle of incidence and crystal position in a block. Without proper normalization, reconstructed images may have artifacts in the form of periodic patterns or other non-uniformities that have been propagated from the data space into the image. Variations in detector response can be quite significant, making normalization an important aspect of the reconstruction process for quantitative accuracy.

The method of Bailey et al. [6] was adapted for the full-ring PET insert system reconstruction [67]. Since the three data types (II, IS, and SS) in the full-ring insert system are circularly symmetrical, and the FOV is entirely inside the insert ring, calculating the components is relatively straightforward. Calculation of the crystal efficiencies, for example, depends on the use of a fan-sum algorithm to process uniform cylinder emission data [4]. This algorithm assumes that the sum of fan-beam measurements for each view angle should be constant among all views. Under this assumption, it becomes possible to find the efficiency for the vertex crystal of each fan. Unfortunately, the circular asymmetry of the half-ring insert invalidates this assumption; the insert attenuation affects each view differently, and the II fan-beam measurements are truncated at every view angle.

We first tried to bring the SS dataset into conformance with the fan-sum assumption by subtracting randoms and scatter and then dividing by the insert attenuation (described in Section 6.4), i.e.,

$$\tilde{d}_{2,i} \triangleq \frac{d_{2,i} - \bar{r}_{2,i} - \bar{s}_{2,i}}{e^{-\int_{L_{2,i}} \mu_{\text{insert}}\, dl}}, \tag{6.5}$$

where the integral in the denominator represents the forward projection through an attenuation map of the insert along the ith SS LOR. Some fan angles pass through a significant amount of LSO, leading to a denominator $\ll 1$. These LORs mainly contain randoms and scattered counts, which have been corrected for in mean, but not on a count-by-count basis; in fact, many of these LORs have negative counts after the subtractions in the numerator. Therefore, the counts in these LORs become over-amplified (in the positive and negative directions), leading to poor normalization factors using the standard fan-sum algorithm.

We determined that a more principled approach for the half-ring insert system is to use a maximum likelihood technique to jointly estimate all normalization components simultaneously. Bai et al. [5] emphasized that by using the same statistical framework for both the image estimation (reconstruction) and the normalization factor estimation, the normalization factor estimates will be consistent with the other factors comprising the data model. In contrast, the more standard component-based methods cited above sequentially calculate the various components using a different method for each component.

Hogg et al. [31] and Bai et al. [5] independently proposed ML-based normalization approaches, although the modeled components were slightly different. They both modeled the crystal efficiencies and geometric effects, which are the main normalization components. The geometric modeling accounts for the angle of incidence that is formed between an LOR and the detectors. It also accounts for crystal interference, which is related to the position of the crystal within a block in the transverse and axial directions [12]. These geometric effects are usually modeled in the diagonal normalization factor as opposed to in P_{geom}. For the purpose of computing the ML normalization factors, the emission image variables are taken to be the known source distribution (e.g., uniform cylinder or rotating rod source), and the normalization factor variables are the only unknowns. The ML solution can be obtained using steepest ascent, coordinate ascent, or any other optimization method.

In our work, we have already accounted for all the geometric effects through the modeling of block detectors in their native geometry in P_{geom}. It is more natural to account for these effects in P_{geom} rather than in the normalization factor, since the physics of intercrystal penetration and crystal interference can actually be calculated directly. Additionally, the geometric normalization effects are usually computed using various types of LOR symmetry in the PET scanner; this type of symmetry is reduced or difficult to identify due to the half-ring insert.

For our system, we express the normalization components that form the diagonal entries of P_{norm} as

$$[P_{\text{norm}}]_{ii} \triangleq \varepsilon_{k^{i1}}\, \varepsilon_{k^{i2}}\, \eta_{m^i}\, \alpha_i, \tag{6.6}$$

where i is the LOR index, k is the absolute crystal index, m is the sinogram index, $\varepsilon_{k^{i1}}$ is the efficiency of crystal 1 in the LOR, $\varepsilon_{k^{i2}}$ is the efficiency of crystal 2 in the LOR,

η_{m^i} is the efficiency of the sinogram which contains the LOR, and α_i is the attenuation due to the insert along the LOR. The coincidence type index has been suppressed for more convenient notation. We will only estimate the crystal and sinogram efficiencies in the ML optimization; α_i is treated as a component of the normalization factor, but is determined beforehand from the forward projection and exponentiation of the insert attenuation map obtained in Section 6.4.

Although η_{m^i} does not typically appear as a component in component-based normalization methods, we include it here as an overall scale factor for each sinogram. It is necessary to include this factor in the model since certain IS sinograms are always zero, due to the wiring of our insert to the PET scanner's coincidence processing hardware; for these sinograms, the crystal efficiencies should still be nonzero (since the same crystals are used in other sinograms too), but the overall sinogram weighting should be zero. This scale factor could also become useful if the fully 3D acquired data is compressed axially prior to reconstruction, in which case the sinograms could have a variable number of ring combinations added together. (Axial compression is an added feature that was not extensively tested or used in this work.)

Let $\boldsymbol{\theta} \triangleq [\boldsymbol{\varepsilon}; \boldsymbol{\eta}]$ represent the parameter vector to be estimated. Our expanded mean data model from (6.1), now as a function of $\boldsymbol{\theta}$ instead of $\boldsymbol{\lambda}$, therefore becomes

$$g_i(\boldsymbol{\theta}) = \varepsilon_{k^{i1}}\, \varepsilon_{k^{i2}}\, \eta_{m^i}\, h_i + \bar{r}_i + \bar{s}_i, \tag{6.7}$$

where

$$\boldsymbol{h} \triangleq \operatorname{diag}\{\boldsymbol{\alpha}\} \cdot P_{\text{atten}} \cdot P_{\text{geom}} \cdot \boldsymbol{\lambda} \tag{6.8}$$

is the forward projection of the known emission image using all factors other than those to be estimated. Note that the scatter estimate is not multiplied by the crystal efficiencies in our current implementation, although it is unlikely to make a significant difference in the final reconstruction. The Poisson likelihood function is then given by

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_i d_i \ln g_i(\boldsymbol{\theta}) - g_i(\boldsymbol{\theta}) && (6.9) \\ &\equiv \sum_i d_i \ln(\varepsilon_{k^{i1}}\, \varepsilon_{k^{i2}}\, \eta_{m^i}\, h_i + \bar{r}_i + \bar{s}_i) - \varepsilon_{k^{i1}}\, \varepsilon_{k^{i2}}\, \eta_{m^i}\, h_i\,, && (6.10) \end{aligned}$$

where '≡' indicates that terms not dependent on $\boldsymbol{\theta}$ were left out of the expression.

We implemented a steepest ascent method (with line search) and the well-known Broyden-Fletcher-Goldfarb-Shanno (BFGS) method to perform the maximization of (6.10). Both methods require the calculation of $\nabla L(\boldsymbol{\theta})$, but not the Hessian (which would be complicated to write down analytically). For completeness, we provide the gradient components here:

$$\frac{\partial L}{\partial \varepsilon_{k'}} = \sum_{\{i:\, k^{i1}=k'\}} \varepsilon_{k^{i2}}\, \eta_{m^i}\, h_i \cdot \left[\frac{d_i}{g_i(\boldsymbol{\theta})} - 1\right] + \sum_{\{i:\, k^{i2}=k'\}} \varepsilon_{k^{i1}}\, \eta_{m^i}\, h_i \cdot \left[\frac{d_i}{g_i(\boldsymbol{\theta})} - 1\right] \tag{6.11}$$

$$\frac{\partial L}{\partial \eta_{m'}} = \sum_{\{i:\, m^{i}=m'\}} \varepsilon_{k^{i1}}\, \varepsilon_{k^{i2}}\, h_i \cdot \left[\frac{d_i}{g_i(\boldsymbol{\theta})} - 1\right] \tag{6.12}$$

The optimization methods are iterative and therefore need to be run to convergence. They also must be initialized with a reasonable guess for the components (e.g., a uniform guess that is approximately the same order of magnitude as the final estimate). One major benefit of the ML method for determining the normalization factors is that negative numbers are not encountered; randoms and scatter are not subtracted from the measured data, but are instead incorporated directly into the complete data model.

6.2.4 Randoms and Scatter Estimation

The mean number of randoms and scattered counts are assumed to be known prior to the reconstruction. An estimate of the randoms is measured during each acquisition using a delayed coincidence timing window. The unprocessed delayed counts data has statistical noise modeled by the Poisson distribution. Therefore, the mean estimate of $\bar{r}_{t,i}$ should ideally include some form of smoothing to prevent the introduction of extraneous noise into the reconstruction. A variance-reduced estimate can be obtained using a maximum-likelihood procedure that first estimates the singles rate at each crystal, and then takes the product of the singles rate at the two crystals forming an LOR [62, 32]. The product of the singles rates, multiplied by the coincidence timing window, gives the smoothed randoms estimate at each LOR. Since the ML methods for estimating the singles rates do not easily apply to the insert system geometry, and other methods are comparably difficult to apply (e.g., [58]), we have forgone the use of variance-reduction techniques at this time; they could lead to bias without careful design and study. By using the noisy delayed counts directly for the randoms estimate, the image update might be noisier than if we had used a smoothed estimate. However, this only affects the variance of the image, and not the mean value.

Unlike randoms, the scattered counts estimate cannot be measured by our system, and must be determined by some other means. Compton scattering is the predominant form of interaction of 511 keV gamma rays in biological tissue. In a Compton interaction, a gamma ray collides with an electron, loses some energy to the electron, and continues in a different path according to the Compton scattering equation. Energy discrimination hardware implemented on a PET scanner is only somewhat able to prevent scattered counts from being recorded, as the energy resolution of detector materials is not good enough to fully distinguish between scattered gamma rays (energy < 511 keV) and unscattered gamma rays (energy $= 511$ keV). Therefore, some other method is needed to account for the scattered coincidences that were recorded. Markiewicz et al. [56] provides a good review of various scatter estimation methods, along with a method to perform accurate scatter estimation by using a multiple scatter model.

Our scatter estimate is currently based on the single scatter simulation (SSS) method [103], in which one gamma ray in the pair is assumed to scatter once. This method uses the analytical Klein-Nishina formula to calculate the differential Compton scattering cross section, $d\sigma/d\Omega$. The scatter geometry as it is applied to the SSS method is shown in Figure 6.3. For simplicity, it has been drawn in 2D, but the actual computations are performed in 3D. For each LOR, an integration is performed over the scattering volume. From Watson [103], the scatter contribution for the LOR defined by detectors A and B is calculated as

$$s_{AB} = k \int_{V_{\text{scatter}}} \frac{\sigma_{AS}\sigma_{BS}}{4\pi R_{AS}^2 R_{BS}^2} \frac{\mu d\sigma}{\sigma d\Omega} [I^{(A)} + I^{(B)}] \, dV, \tag{6.13}$$

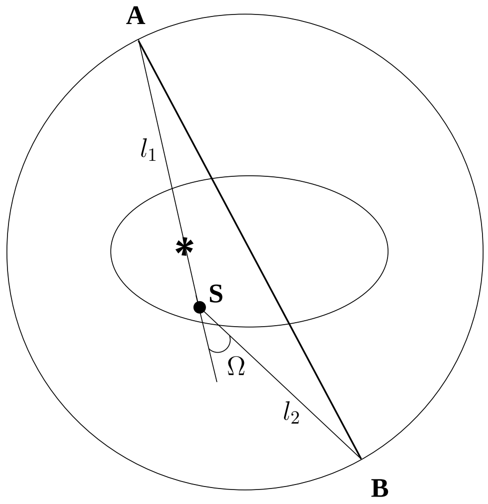


Figure 6.3: For the LOR defined by detectors A and B, the SSS method calculates the scatter contribution from each possible scatter point (S) in the object volume. The asterisk (*) represents a hypothetical annihilation location. The LOR and scatter point together define the scatter angle Ω at which the Klein-Nishina formula is evaluated.

where

$$I^{(A)} = \varepsilon_{AS}(E)\varepsilon_{BS}(E')\exp\left[-\int_A^S \mu(E,l_1)\,dl_1 - \int_S^B \mu(E',l_2)\,dl_2\right]\int_S^A \lambda(l_1)\,dl_1 \tag{6.14}$$

$$I^{(B)} = \varepsilon_{AS}(E')\varepsilon_{BS}(E)\exp\left[-\int_A^S \mu(E',l_1)\,dl_1 - \int_S^B \mu(E,l_2)\,dl_2\right]\int_S^B \lambda(l_2)\,dl_2. \tag{6.15}$$

In these equations, k is a scaling coefficient, σ_{AS} and σ_{BS} are the detector cross sections seen by the scatter point, R_{AS} and R_{BS} are the lengths between the detectors and scatter point, μ is the spatially-variant and energy-dependent attenuation coefficient, σ is the total Compton scattering cross section, and finally, ε_{AS} and ε_{BS} are the detection efficiencies at $E = 511$ keV or E', the scattered photon energy. E' can be determined from the Compton scattering equation,

$$E'(\Omega) = \frac{E}{1 + \frac{E}{mc^2}(1 - \cos\Omega)}, \tag{6.16}$$

where m is the resting mass of the electron prior to scattering, c is the speed of light, and Ω is the scatter angle corresponding to the current scatter point.

The body attenuation map has already been determined from a CT scan, as explained in Section 6.2.2. Since the SSS line integral calculations through the μ map are energy dependent, the body attenuation coefficients are scaled using the same approach as for the conversion from CT energies to 511 keV. However, the insert attenuation map cannot be readily scaled to other energies without knowledge of the material composition at each voxel. Although the insert device contains other materials in addition to LSO (e.g., the aluminum casing, PMTs, and electronics), we focus on the LSO, since it is the most highly attenuating medium within the insert. In the SSS calculations, we analytically compute the line integral through the insert LSO crystals, using their known geometry and energy-dependent attenuation coefficients. The body attenuation map is discretely forward projected. These calculations are performed from the scatter point to each detector in the LOR.

The emission activity distribution required by the SSS method is based on an initial reconstruction in which scatter has not been corrected. The scatter estimate can be improved after every few iterations of the EM algorithm (e.g., by using new estimates for $\boldsymbol{\lambda}$), but this was not considered in the present work; it was assumed that the preliminary scatter-uncorrected reconstruction was sufficiently accurate for the final scatter estimation.

MC simulations have shown that few coincidence events are detected when one or both of the gamma rays have been scattered by LSO in a different detector block. This obviates the need to include the insert crystals in the scattering volume integral calculations, as the probability of detection after scattering in the insert is negligibly small. The simulations also indicate that the single-scatter model is appropriate, as the number of multiply-scattered events that were detected is much smaller than the number of detected events in which there was only one Compton scatter interaction.

Finally, it is necessary to scale the scatter estimate to the measured data via the scale factor, k. To find an appropriate scaling, we first compress the fully 3D sinograms axially into a 2D sinogram; this is done separately for the prompts data, the delayed counts data, and the scatter estimate. Since the measured prompts data consists of trues, randoms, and scattered counts, the delayed counts sinogram is subtracted from the prompts data to remove the randoms. The resulting sinogram will then approximately contain only trues and scattered counts. Furthermore, the outer fan portion of this sinogram will contain only scattered counts (assuming the object is smaller than the FOV). A constant scale factor for the scatter estimate is finally determined from these scatter-only tail regions of the sinogram. This procedure is done separately for the II, IS, and SS data.

6.3 Symmetry and Parallelization Approach

As in the helical CT work, we use lookup tables to store the precomputed system matrix for fast access. The size of the fully 3D system matrix lookup tables can be reduced significantly through symmetry considerations. Due to the half-ring insert geometry, only two-fold mirror symmetry is possible for II and IS in the transverse direction; eight-fold rotational and mirror symmetry can be exploited for SS since the attenuation due to the half-ring insert is not factored into the spatially variant system matrix weights (i.e., P_{geom}), but is instead treated as part of the diagonal normalization matrix.

In the axial direction, mirror and translational symmetry exist for all three coincidence types. Like above, SS translational symmetry is possible since the attenuation due to the insert is treated separately from the spatially variant weights. To obtain maximum axial symmetry, the positioning and thickness of the slices needs to be carefully considered. Figure 6.4 shows the placement of the slices relative to the crystals. The slices were designed to have the same thickness as the insert crystal pitch, and for their edges to line up with the insert crystal edges. The slices line up with the center or adjacent edges of the scanner crystals, which are twice as thick. We approximate the gap between axial blocks by one extra crystal ring to preserve translational symmetry; the crystal ring and gap dimensions are approximately the same for both the insert and scanner crystals, respectively.

As an example, Figure 6.5 shows the axial base symmetry LORs and symmetrically related LORs for the IS type. The two thick solid lines are the base symmetry LORs, and the thinner dotted lines represent the symmetrically related LORs. There are two base symmetry LORs since the IS symmetry only applies to every other insert ring.

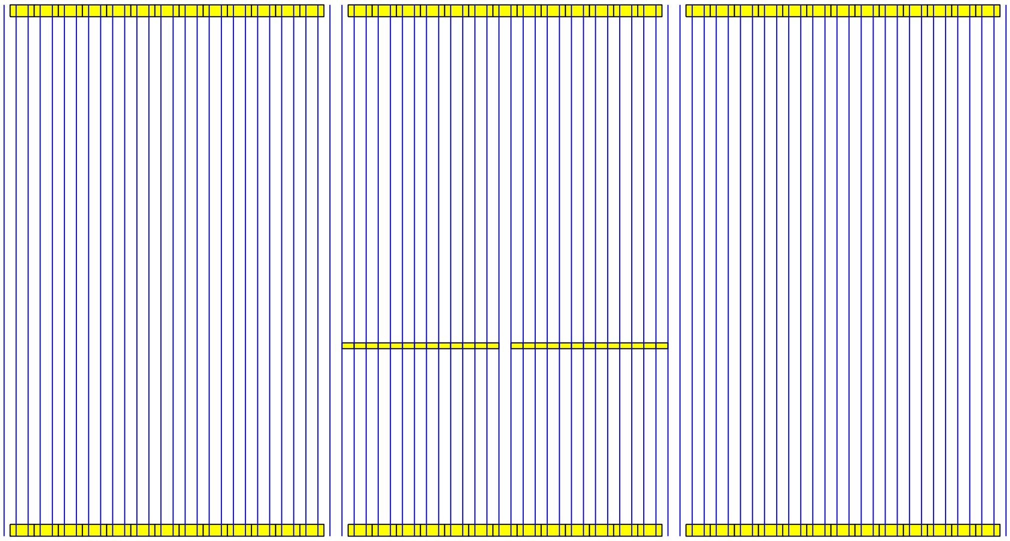


Figure 6.4: Axial view of insert system showing image slices in relation to the insert and scanner crystals. Slices are shown as blue vertical lines, and each ring of crystals is shown as a yellow block.

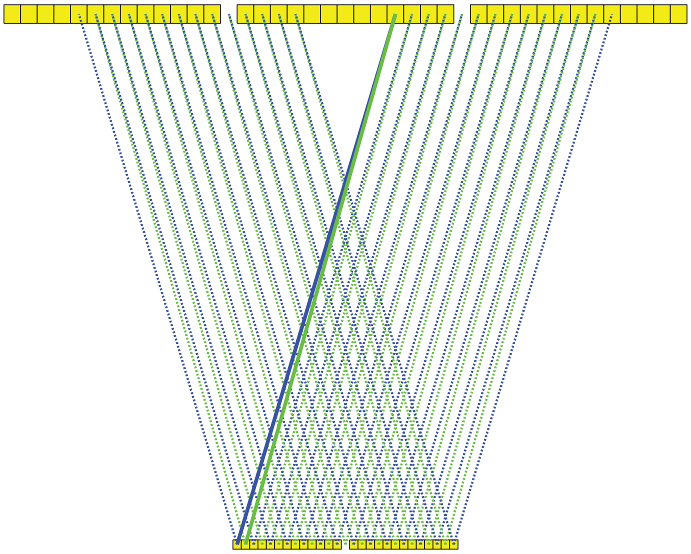


Figure 6.5: Axial symmetry for the IS coincidence type shown for a particular ring difference. Scanner rings are shown on top, and insert rings on bottom.

We use three Michelograms (one for each coincidence type) to specify the ring combinations to use in the reconstruction. A Michelogram is a table that conveniently indexes the particular ring combinations to be used in the reconstruction; one axis corresponds to the first ring index in an LOR, and the second axis corresponds to the second ring index in an LOR [23]. The maximum ring difference is a user-specified value in our implementation that provides flexibility over how many computations are performed in the forward and backprojections. Additionally, axial compression was partially implemented, but will need further work. Adjusting the axial compression is more difficult than adjusting the maximum ring difference, since the normalization factors are typically determined using the uncompressed data. Using all ring differences without axial compression, the maximum number of 2D sinograms is 378 for II, 1107 for IS, and 861 for SS.

Based on the transverse/axial symmetry considerations and the structure of the Michelograms, we have parallelized the forward and backprojections to fully utilize the available computational resources on a multi-core workstation. As in the CT chapter, the symmetry is inherently connected with the parallelization approach because it determines the granularity of the parallelization. The parallelization performed here is done at the granularity of a 2D sinogram, which corresponds to the LORs in one ring combination. Each 2D sinogram array can be thought of as a single entry in the Michelogram. Since this is a finer level of granularity than that of the helical CT reconstruction algorithm, we use the dynamic scheduling capabilities of OpenMP to assign the forward and backprojections to the threads as they become available. The amount of time to process each 2D sinogram (regardless of whether it is for forward or backprojection) is approximately the same. This ensures the threads are well-balanced. The scheduling is done in such a way that for each lookup table that is read into memory, all symmetric 2D sinograms will be processed at the same time to minimize LUT file reading time and to maximize cache performance.

6.4 Prototype Insert System

Figure 6.6 shows the front and back views of the insert positioned in the gantry of the clinical PET/CT scanner. The insert (without its cover) is shown in Figure 6.7(a), and a single detector block is shown in Figure 6.7(b). Each detector module consists of a 13×13 array of LSO crystals, light guide, position-sensitive PMT (to convert the scintillation light into measurable electrical current), and readout electronics. The system configuration is described in more detail in Wu et al. [106].

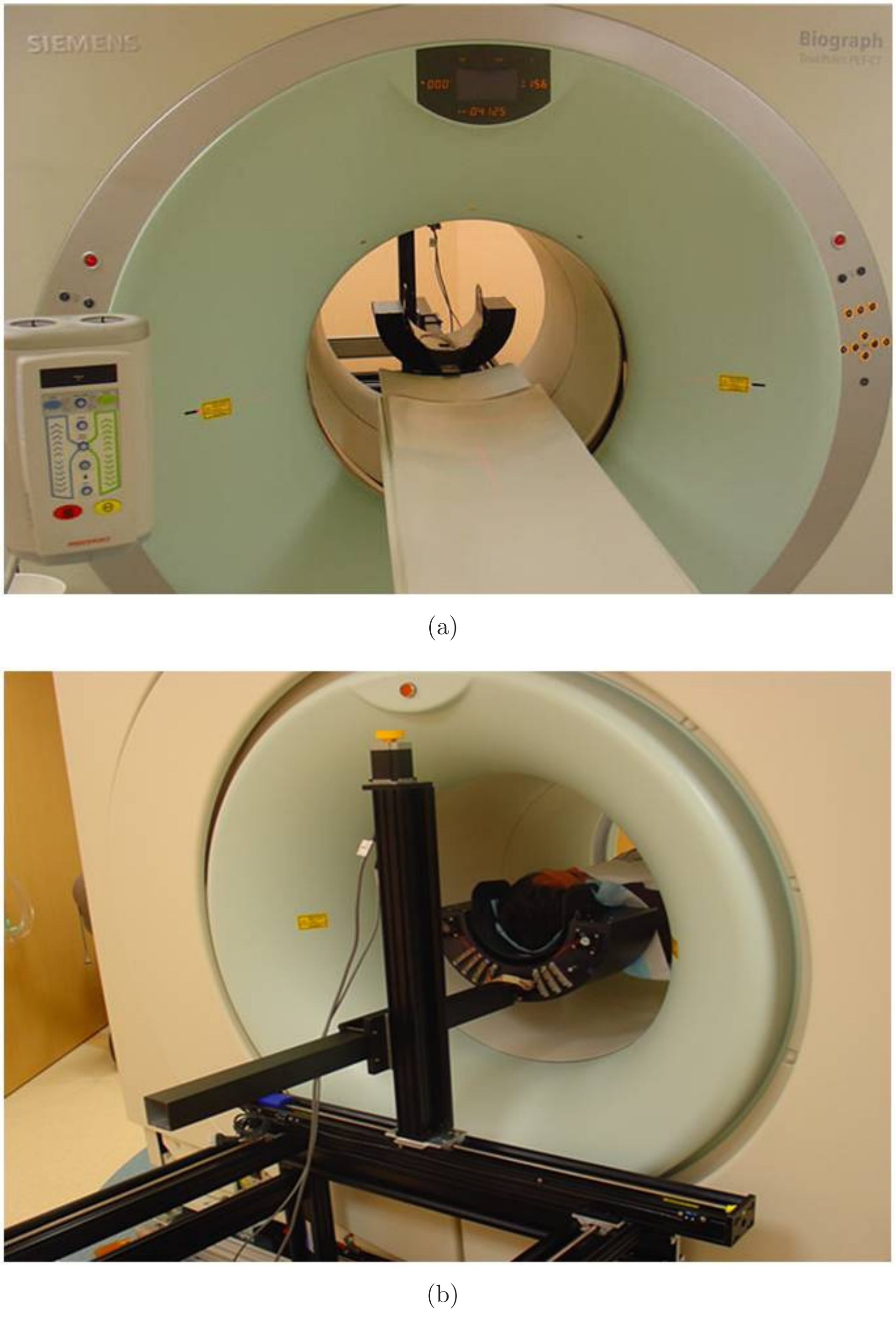
To determine the α_i factors in the component-based normalization model, it is necessary to obtain an accurate attenuation map of the insert system. The LSO crystals are very dense, as they must be able to stop 511 keV gamma rays. This unfortunately makes it impossible to scan the insert device using a medical CT scanner like we are able to do for the body; the maximum energy of the x-ray photons emitted from the x-ray tube is about 120 keV, which is far less than the amount of energy needed to penetrate the LSO crystals for transmission tomography. There are also concerns that the large number of absorbed x-ray photons could damage the PMTs and detector module electronics.

For these reasons, we set up a rotation stage to use positron-emitting line sources as a transmission source, and the scanner crystals operating in coincidence mode as the detectors. The use of a positron source instead of x-rays makes it unnecessary to scale the attenuation coefficients from one energy level to another. However, the rotation stage was completely independent of the scanner, therefore precluding the use of sinogram windowing to reduce the scatter contribution to the transmission scan. A long blank scan and transmission scan of the insert (each about 24 hours) were acquired. Single-slice rebinning was performed on both datasets, which were subsequently reconstructed using a 2D version of the AM algorithm. One slice of the reconstructed image volume is shown in Figure 6.8. Due to the inclusion of scatter, the entire reconstruction FOV looks noisy. More importantly, the LSO attenuation coefficients were more than five times smaller than expected , and voxels that should have the same value (e.g., all LSO voxels) were actually highly non-uniform. This attenuation map is clearly not accurate enough to use in the calculation of the α_i factors.

To understand the cause of these discrepancies, we performed a series of analytical and MC simulations that used the same acquisition geometry as the real transmission scan (albeit in 2D). Only the attenuation from the LSO crystals was simulated, as

---

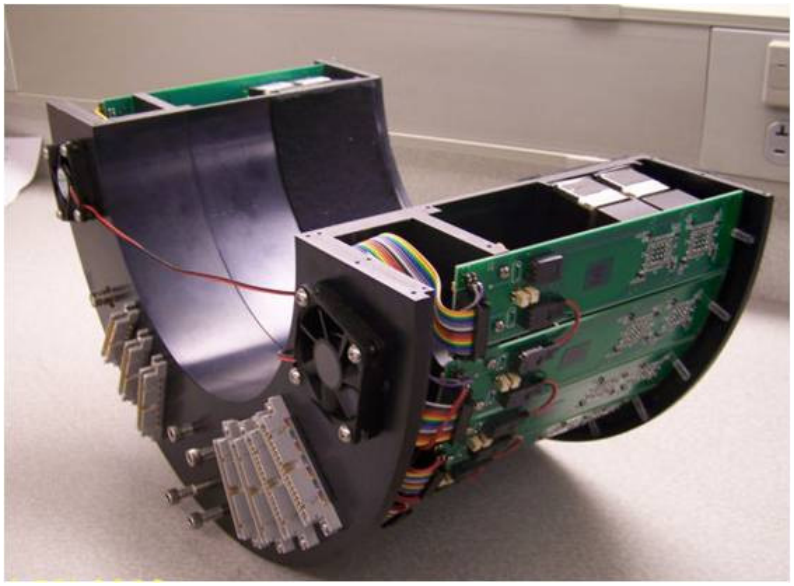
LSO has a linear attenuation coefficient of 0.087 mm^{-1} at 511 keV [54, p. R291].



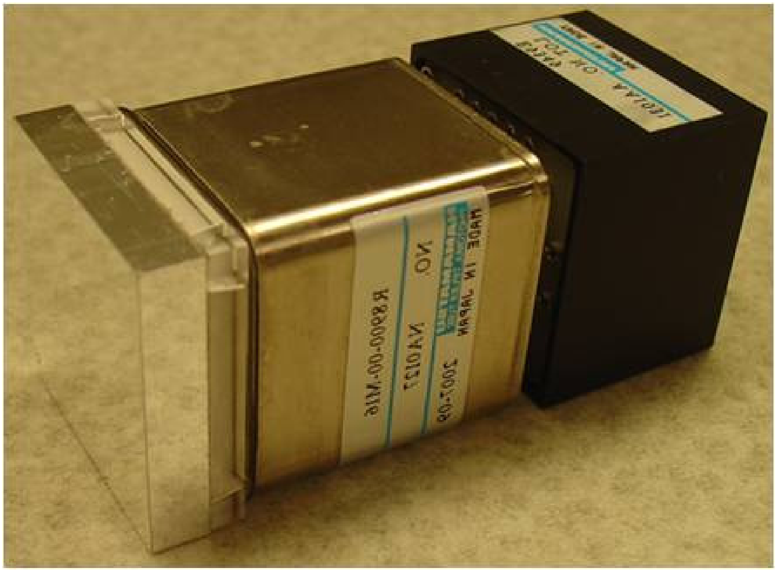
(a)

(b)

Figure 6.6: Siemens Biograph 16 PET/CT scanner with insert in FOV. (a) View from front of scanner; (b) view from back of scanner, with subject's head centered in the insert FOV (for demonstration purposes only).



(a)



(b)

Figure 6.7: Close-up photographs of the insert system. (a) The insert gantry with its cover removed; (b) an individual detector block consisting of a 13×13 array of LSO crystals, light guide, position-sensitive PMT, and electronics. (Reprinted by permission of the IEEE from [106]. ©2008 IEEE.)

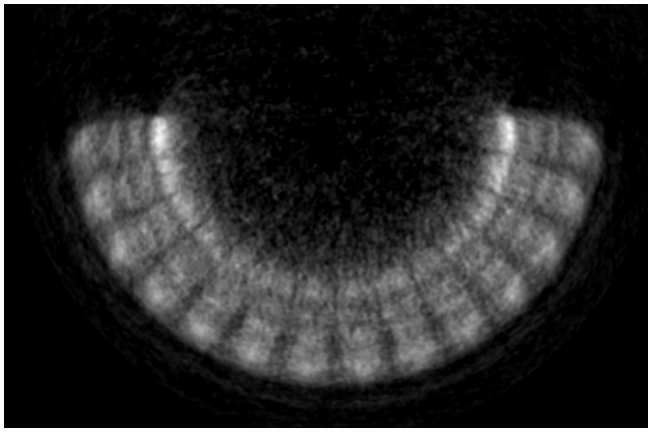
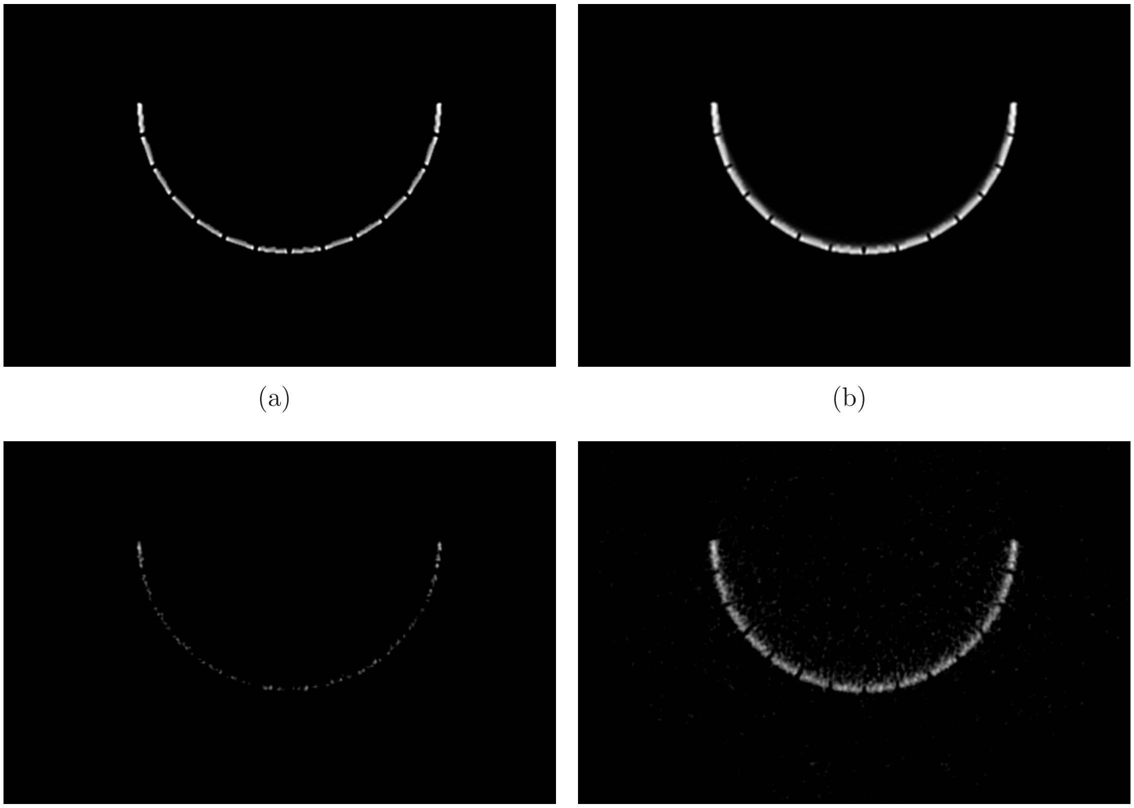


Figure 6.8: Insert transmission image from the rotating line source scan. The LSO crystals form the innermost arc in the image. Window min/max = [0.0, 0.016] mm^{-1} (full window display).

the crystals have a higher attenuation coefficient than the insert's aluminum housing and the other components of the detector module. The simulations were performed without saving randoms or scattered coincidences so as to determine whether a strictly narrow-beam acquisition could improve the quantitative accuracy of the reconstructed image. Figures 6.9(a) and 6.9(b) show the reconstruction of the LSO half-ring from analytically-generated data, and Figures 6.9(c) and 6.9(d) show the reconstruction of the LSO half-ring from the MC simulation. The left column uses a narrow image window, while the right column uses the full window. It can be observed in these images that the uppermost LSO crystals have higher attenuation coefficients than the other crystals. This indicates that non-uniformity among the LSO voxels is still present, even without randoms or scatter (or noise, in the case of the analytical simulation).

Given these results, it was then hypothesized that the problem was caused during the transmission reconstruction by invalid assumptions made in the linear forward projection and subsequent exponentiation in the LORs passing through a long path of LSO. Since the insert LSO crystals are only 5 mm thick, and the scanner crystals (being used as transmission detectors) have a 4×4 mm front face, individual LORs may detect activity passing tangentially through the LSO but also through non-LSO



(a) (b)

(c) (d)

Figure 6.9: Reconstructions of the simulated LSO half-ring transmission scans. The attenuation coefficients were smaller than expected, and were non-uniform over the crystals in the half-ring. Top row: analytical simulation, bottom row: MC simulation. Left column window min/max = [0.04, 0.09] mm^{-1}, right column window min/max = [0.0, 0.09] mm^{-1}.

material on either side. These LORs experience a large variation in the measured signal over the spatial extent of the LOR, i.e., they have not been sufficiently sampled.

Mathematically, this can be understood by examining the derivation of the forward projector in Politte and Whiting [72], which approximates the exponential function by the Maclaurin series expansion to first order. For this linear approximation to hold, the argument of the exponential must be much smaller than 1. Specifically, the argument is the *variation* in the line integrals through the attenuation map, taken over the face of a detector (or in our case with coincidence detection, over both detector faces comprising an LOR). The variation will be large if certain sub-rays of a given LOR pass through a long path of LSO, while others do not.

We ran another analytical simulation to show that the dense LSO crystals are indeed responsible for the problems with the reconstructed insert attenuation map. Instead of specifying the half-ring crystal material as LSO in the simulation, we specified it as water (which has an attenuation coefficient almost an order of magnitude less than LSO). Everything else in the simulation remained exactly the same; the only change was to the insert crystal material.

In Figure 6.10, we show the reconstruction of this simulated data using a narrow and wide image window. There is no non-uniformity in the crystal values, and the attenuation coefficients are quantitatively correct. This leads to the conclusion that attenuation along LORs passing tangentially through the insert cannot be accurately modeled using our current projection operator.

Not only does this have implications on the manner in which we form an attenuation map of the insert, it also affects the forward projection of the insert attenuation map (once we have one) for the purpose of calculating the α_i factors accurately. It implies that a diagonal attenuation matrix factor for the insert cannot adequately model these particular LORs. For the time being, we have therefore chosen to restrict the maximum diameter of objects being scanned to ensure that their fan-width is smaller than the fan-width of the insert half-ring crystals (as seen by the SS coincidence type). While it might be possible to incorporate the insert attenuation calculations

---

A more physically meaningful term that some authors use instead of "LOR" is "tube of response"; the tube of response is the finite volume that connects the two crystals forming a detector pair.


(a) (b)

Figure 6.10: Reconstructions of the analytically-simulated water half-ring transmission scan. (a) Window min/max = [0.0075, 0.0115] mm^{-1}; (b) window min/max = [0.0, 0.012] mm^{-1}.

on a sub-ray basis into the P_{geom} factor, this would be computationally expensive in fully 3D. It was done in 2D in Chapter 5, but we have not investigated it in the 3D work.

To circumvent this fundamental problem, we performed a transmission scan of a single detector block using a microPET Focus 220 small-animal scanner (CTI/Concorde Microsystems, Knoxville, TN). A ^{68}Ge positron point source was rotated helically through the scanner FOV to acquire a high-statistics 44-hour blank scan and 60-hour transmission scan. By scanning only one block instead of the full insert gantry, we were able to avoid acquiring LORs that pass through a large amount of LSO material. Since the rotation stage is a built-in feature of the microPET scanner, we were able to use its sinogram windowing feature to reduce the number of recorded scattered counts.

The scan was reconstructed using a MAP algorithm on the microPET console. From the small reconstructed image of the detector block and a CAD drawing of the insert gantry, we manually generated a composite 3D attenuation map of the entire insert device. To do this, we used image processing techniques to position and rotate the detector image into the 28 exact block locations in the attenuation map. The dimensions and shapes of the metal structures supporting the detectors were determined from the CAD drawing, and were subsequently voxelized into the correct positions in

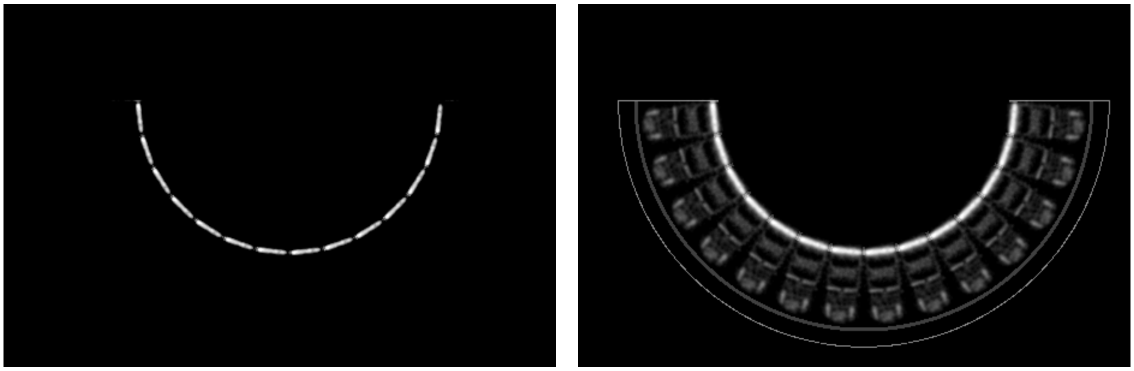

(a) (b)

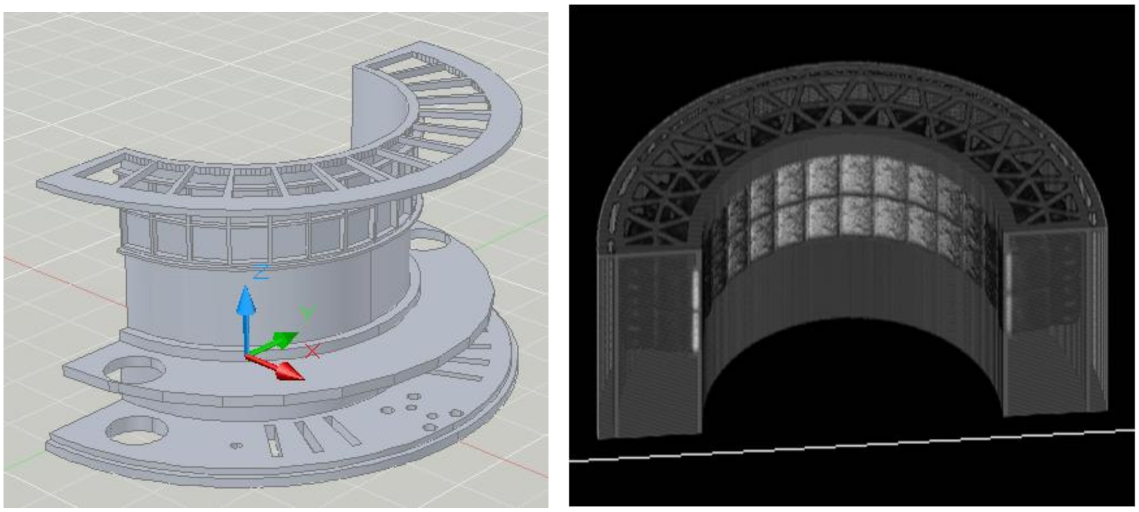
Figure 6.11: One slice of the composite insert attenuation map. (a) Window min/max = [0.04, 0.09] mm^{-1}; (b) window min/max = [0.0, 0.09] mm^{-1}.

the attenuation map. We used our knowledge of the metal composition to approximate the attenuation coefficients at 511 keV for all nonzero voxels other than those containing the detector modules. The front cover design had been changed since the CAD drawing was last updated, so we performed a CT scan of the front cover and segmented it for inclusion in the composite map.

We show a slice of the composite insert attenuation map in Figure 6.11 using a narrow and wide image window. By design, the LSO voxel values are uniform in all detector blocks, and due to the high-quality transmission scan on the microPET, they are close to the true narrow-beam attenuation coefficient for LSO at 511 keV. A rendering of the CAD drawing (with the original front cover pattern) is shown in Figure 6.12(a), and a rendering of the composite attenuation map (with the actual front cover pattern) is shown in Figure 6.12(b).

6.5 Experiments

To test the basic functionality of the insert system, we first positioned a ^{68}Ge line source in the gantry so it was parallel to the axial direction. The line source was scanned in eight different locations within the II and IS FOV. We then sorted the list-mode data into II, IS, and SS detector matrices, and summed the matrices of each



(a)

(b)

Figure 6.12: (a) CAD drawing of insert gantry (without detectors or cover); (b) rendering of the 3D voxelized insert attenuation map.

coincidence type together. The sorting was done without regard to the detector rings in which the coincidence was registered, thereby making this first experiment a compressed 2D dataset. EM reconstructions were performed using different combinations of coincidence types to demonstrate the higher resolution of the insert compared to the scanner crystals. Since the line sources do not produce a large randoms or scatter distribution, we omitted the estimation of these background counts. Furthermore, P_{atten} and P_{norm} were set to the identity matrix for this initial reconstruction experiment.

Next, we acquired a fully 3D dataset to be used for the direct normalization approach. We set up the normalization phantom shown in Figure 6.2, which consisted of two rotating line sources (each with about 760 μCi activity). One line source was placed at a radius of 33.5 cm, and the other was positioned at a radius of 11.5 cm. The rotating line source data was collected for about 30 hours. An image of the normalization phantom was generated based on the source distribution, taking into account the fact that the source moving at a smaller radius spent 33.5/11.5 times longer at each

---

The sinograms that are used in the reconstruction are obtained by converting the detector matrices (with absolute detector indices) into matrices that use relative detector indexing within each fan-beam projection.

position than the source moving at a larger radius. This image was then forward-projected using just the P_{geom} factor, as P_{atten} is approximately equal to the identity matrix for line sources. For the calculation of P_{geom}, the insert and scanner crystals were subdivided four times in-plane, two times axially, and six times in the depth direction.

The direct normalization approach was then tested by scanning a 21.6 cm diameter cylindrical Jaszczak phantom (Data Spectrum Corporation, Hillsborough, NC). One module of this phantom contained cold rods of various diameters surrounded by a warm background. We performed a fully 3D reconstruction on this data using the P_{norm} factor from the rotating line source scan, along with the P_{atten} factor that was estimated from a CT scan of the phantom. Scatter was not estimated since the SSS code was not available at the time the reconstruction was performed. We used all possible ring combinations for II and IS; for SS, we used a maximum ring difference of 10 with span of 1. For comparison with this reconstruction, we performed another scan of the same phantom without the insert in the FOV. This SS data was then reconstructed using our same code, but without normalization; a separate normalization scan had not been acquired for the standalone scanner.

To perform the ML component-based normalization method, we scanned a 20 cm diameter uniform ^{68}Ge cylinder for 18 hours with the insert in the FOV. We compared a steepest ascent method (with secant method for line-search) to a low-memory BFGS implementation [65] to estimate P_{norm}.

Finally, to test this ML normalization approach, we scanned a 9.5 cm diameter cylindrical phantom containing a number of small, low-contrast spheres filled with an ^{18}F solution that resembled tumor activity. The tumor-to-background ratio was 2:1 for these spheres, and they were 0.06, 0.1, and 0.5 mL in volume. This experiment represents a challenging but realistic situation in which there is not much contrast between the tumor and background, and the tumors are very small. A fully 3D reconstruction was then performed using the ML-based P_{norm} and the CT-based P_{atten} factors. We again used all possible ring combinations for II and IS, and for SS, we used a maximum ring difference of 10 with span of 1. A separate reconstruction was performed

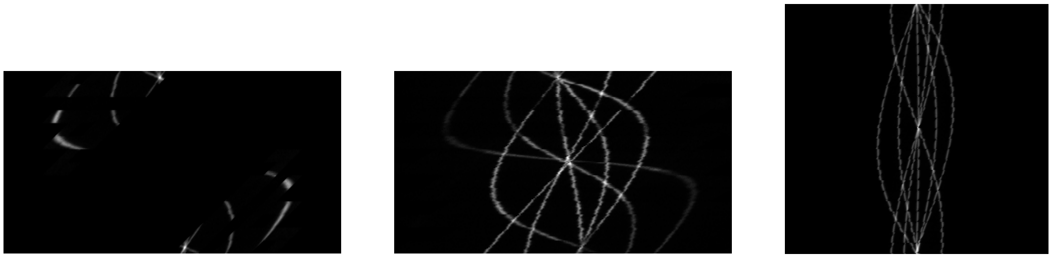

(a) (b) (c)

Figure 6.13: Line source sinograms. (a) II; (b) IS; (c) SS (scaled down and cropped to show only the upper half of view angles).

using only the II and IS data, but initialized with the II+IS+SS reconstruction results. Although somewhat heuristic, we wanted to determine if including the SS data in the joint reconstruction adversely affected image resolution in the insert region.

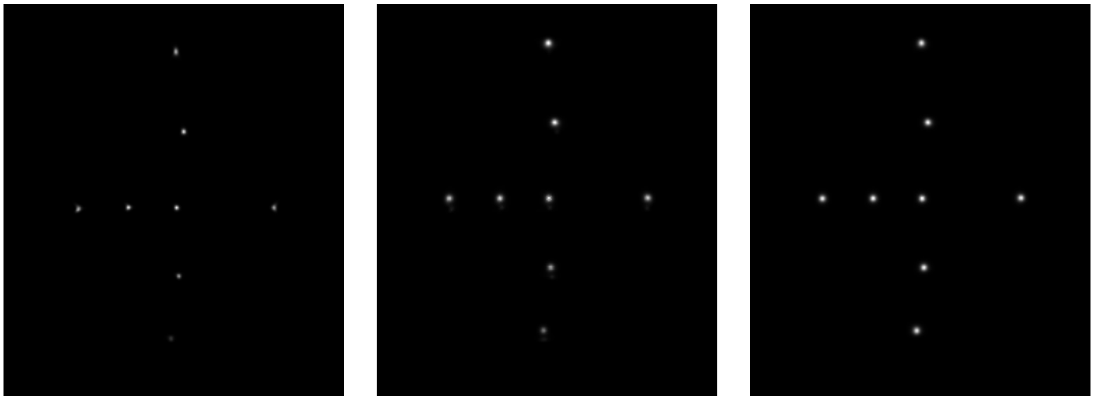
6.6 Results

6.6.1 Line Sources

Figure 6.13 shows the sinograms for the summed line source scans, and Figure 6.14 shows the resulting reconstructions. For reference, the horizontal row of points is just below the upper edge of the II FOV. The reconstructions that incorporate the IS data are able to resolve the line sources better than the SS-only reconstruction, therefore confirming the expected behavior of the insert system and reconstruction algorithm for this simplistic source distribution.

6.6.2 Rotating Line Sources

The sinograms in Figure 6.15 show a comparison of the measured normalization data (left column) to the forward projection of the normalization phantom image using our system model (right column). It should be noted that while the measured and predicted sinograms for each coincidence type look similar, there will still be some differences; the ratio of these two sinograms is used as the normalization factor to



(a) (b) (c)

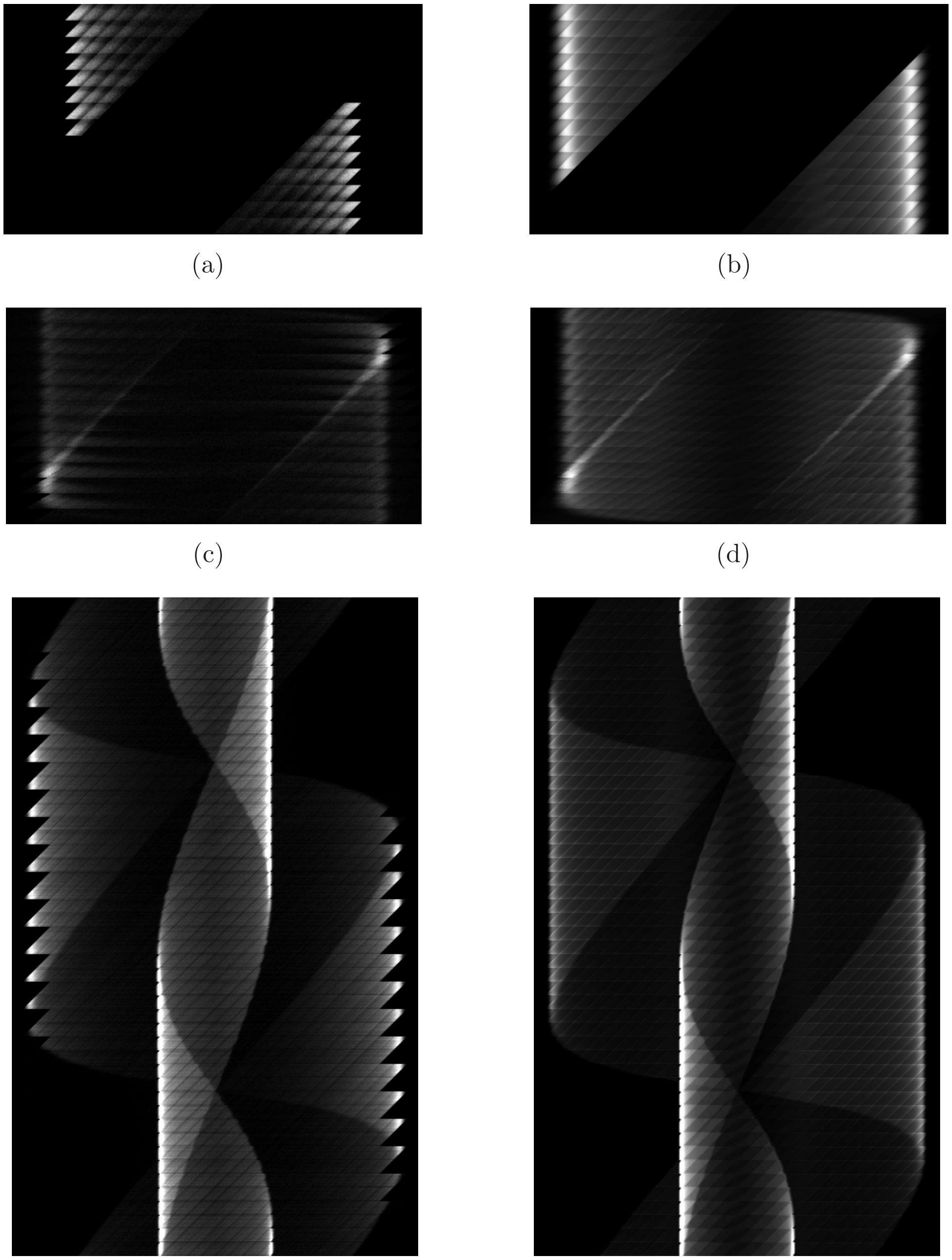
Figure 6.14: Line source reconstructions after 10 EM iterations. (a) IS; (b) II+IS+SS; (c) SS.

account for these differences. Since the line sources were located so close to the edge of the insert and scanner FOVs, and since the scanner had been modified to accommodate the insert electronics, the coincidence processing hardware is not able to detect events in the outermost fan angles. This leads to the jagged appearance seen at the edges of the measured normalization sinograms. The measured II sinogram is most affected, as seen by the truncation of the fan relative to the forward projection sinogram. This effect is automatically accounted for in the direct normalization factor. It is also accounted for in the component-based approach by setting $g_i(\boldsymbol{\theta}) = 0$ for the LORs that are outside the coincidence detection region.

6.6.3 Jaszczak Phantom

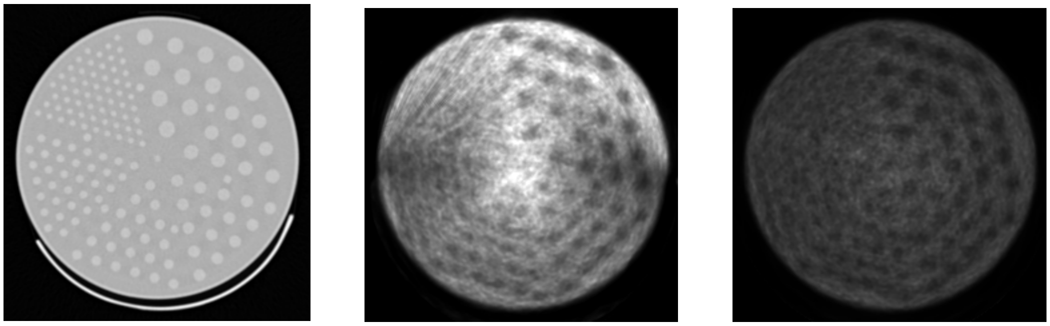
In Figure 6.16(a), we show a CT slice of the Jaszczak phantom to indicate the location and size of the cold rods. Figure 6.16(b) shows the image reconstructed using data from all three coincidence types (i.e., using the data acquired when the insert was in the FOV). Figure 6.16(c) shows the image reconstructed using data from the standalone scanner.

In the CT image, the cold rods show up brighter than the background, and in the PET images, they show up as dark circles. However, there appear to be some significant non-uniformity artifacts in the II+IS+SS image that we will discuss later in this



(a) (b) (c) (d) (e) (f)

Figure 6.15: Comparison of sinograms for the rotating line source phantom for direct normalization. Top row: II, middle row: IS, bottom row: SS. Left column: measured data, right column: forward projection of normalization phantom image.



(a) (b) (c)

Figure 6.16: Reconstructions of the Jaszczak phantom. (a) CT image showing the location and size of the cold rods; (b) reconstruction after 20 EM iterations using all three coincidence types of data (direct normalization applied); (c) reconstruction after 20 EM iterations using standalone scanner data (no normalization applied).

chapter. The standalone image does not suffer from such artifacts, but it too cannot resolve all the cold rods in the two sectors containing the smallest rods.

6.6.4 Uniform Cylinder

The SS, IS, and II sinograms pertaining to the uniform cylinder acquisition are shown in Figures 6.17, 6.18, and 6.19, respectively. Sinogram (a) in the figures is the measured prompt data, sinogram (b) is the estimated prompt data with component-based normalization applied, and sinogram (c) shows the component-based normalization factor itself. A profile through the indicated row in sinograms (a) and (b) is plotted in subfigure (d) for the three coincidence types. The mean prompts estimate is the normalized forward projection of the true cylinder image, plus the randoms and scatter estimates. As can be seen in Figure 6.17(a)-(c), the small bright lines in the central fan that are sloped at a steeper angle than the crystal diagonals correspond to the gaps between the crystal blocks in the insert.

Note that for a normal standalone PET scanner, the sinogram for a centered uniform cylinder would be constant across the view angles and symmetrical within each fan-beam projection (neglecting the variation in crystal efficiencies). This is clearly not

---

Although the measurements and normalization factor estimation were performed in fully 3D, these figures show a single 2D sinogram that was extracted from the fully 3D sinogram.

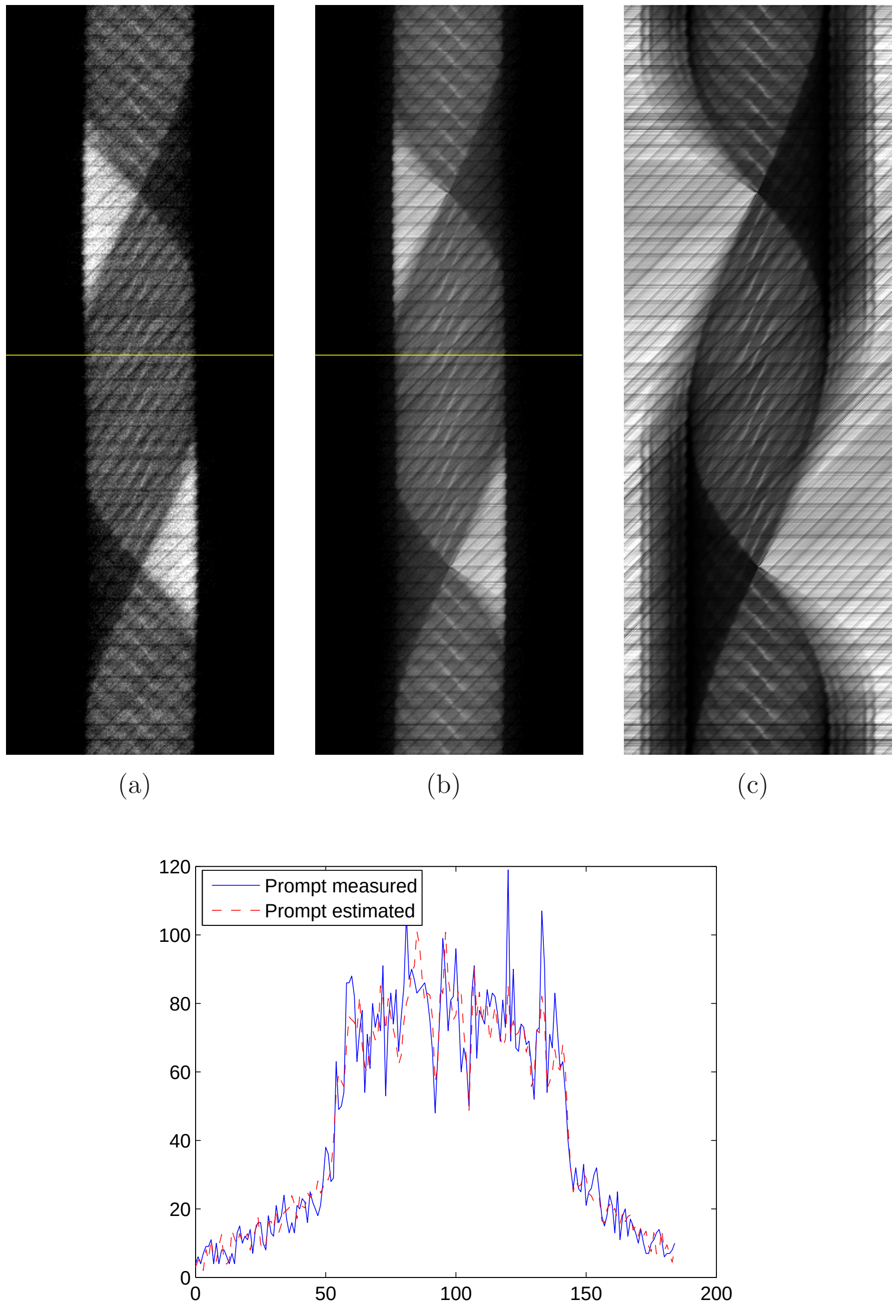
the case when the insert is in the FOV, as shown by the sinograms. Despite the fact that the insert attenuation introduces significant asymmetry into the sinograms, our component-based normalization method models the attenuation correctly and leads to excellent agreement for the SS and IS coincidence types. The II data was somewhat more difficult to model, as seen by the profiles in Figure 6.19(d). This is due to the difficulty of scaling the II scatter estimate to the measured data; the cylinder is so large in the II FOV that there is no scatter-only tail. For the time being, it has to be scaled very approximately, but we discuss an alternative solution to this problem in Chapter 7.

In Figure 6.20, we plot the likelihood function (6.10) versus iteration number to compare the convergence rate of the steepest ascent implementation to the L-BFGS implementation for normalization factor estimation. The L-BFGS method converges significantly faster, and in practice, each iteration is actually faster to compute because it uses a more efficient line-search method (More-Thuente line-search) than the secant method. All normalization factors in this work were therefore computed with L-BFGS.

6.6.5 Small Tumor Phantom

Figure 6.21(a) shows the CT slice through the center of the tumors. Using a very narrow image display window, the plastic spherical shells holding tumor activity can be seen. In Figure 6.21(b), the PET reconstruction of the phantom is shown, using data that was acquired with all four scanner rings without the insert in place. This reconstruction was performed using the built-in reconstruction software on the Siemens Biograph console.

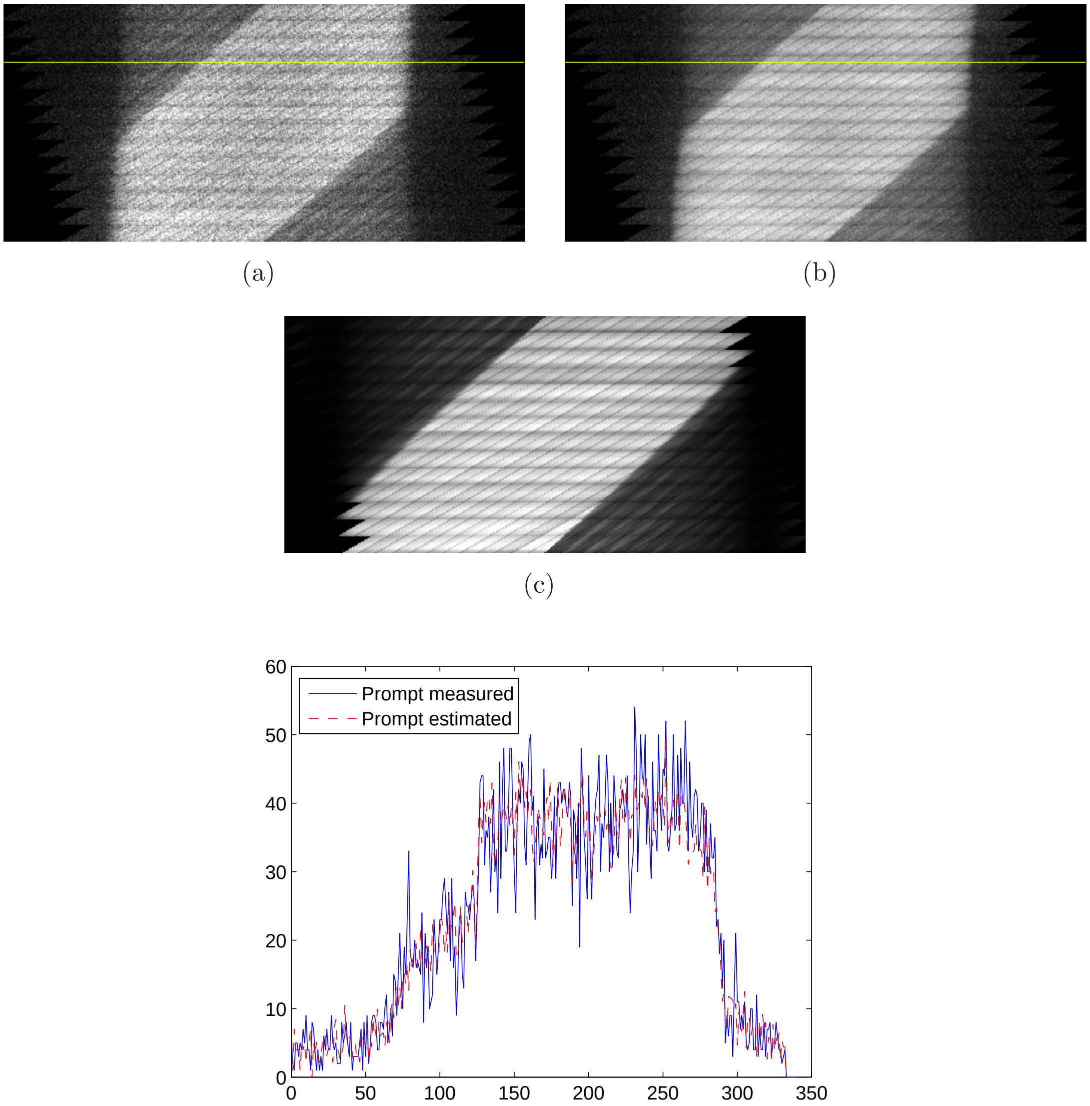
The results from our reconstruction implementation (with insert in the FOV) are shown in Figure 6.22. The image on the left used only the SS data for the reconstruction. The middle image was reconstructed using all three coincidence types, and is shown after performing 25 EM iterations. Finally, the image on the right was reconstructed using the II and IS types after 20 EM iterations. However, to avoid incomplete data artifacts in the rightmost image, the reconstruction was initialized with the image in Figure 6.22(b). All three of these reconstructions made use of the



(a) (b) (c)

(d)

Figure 6.17: SS sinograms and profile for the uniform cylinder acquisition for component-based normalization. (a) measured prompt sinogram; (b) estimated mean prompt sinogram; (c) ML normalization factor; (d) profile through the indicated row in sinograms (a) and (b). The image window is the same for (a) and (b).

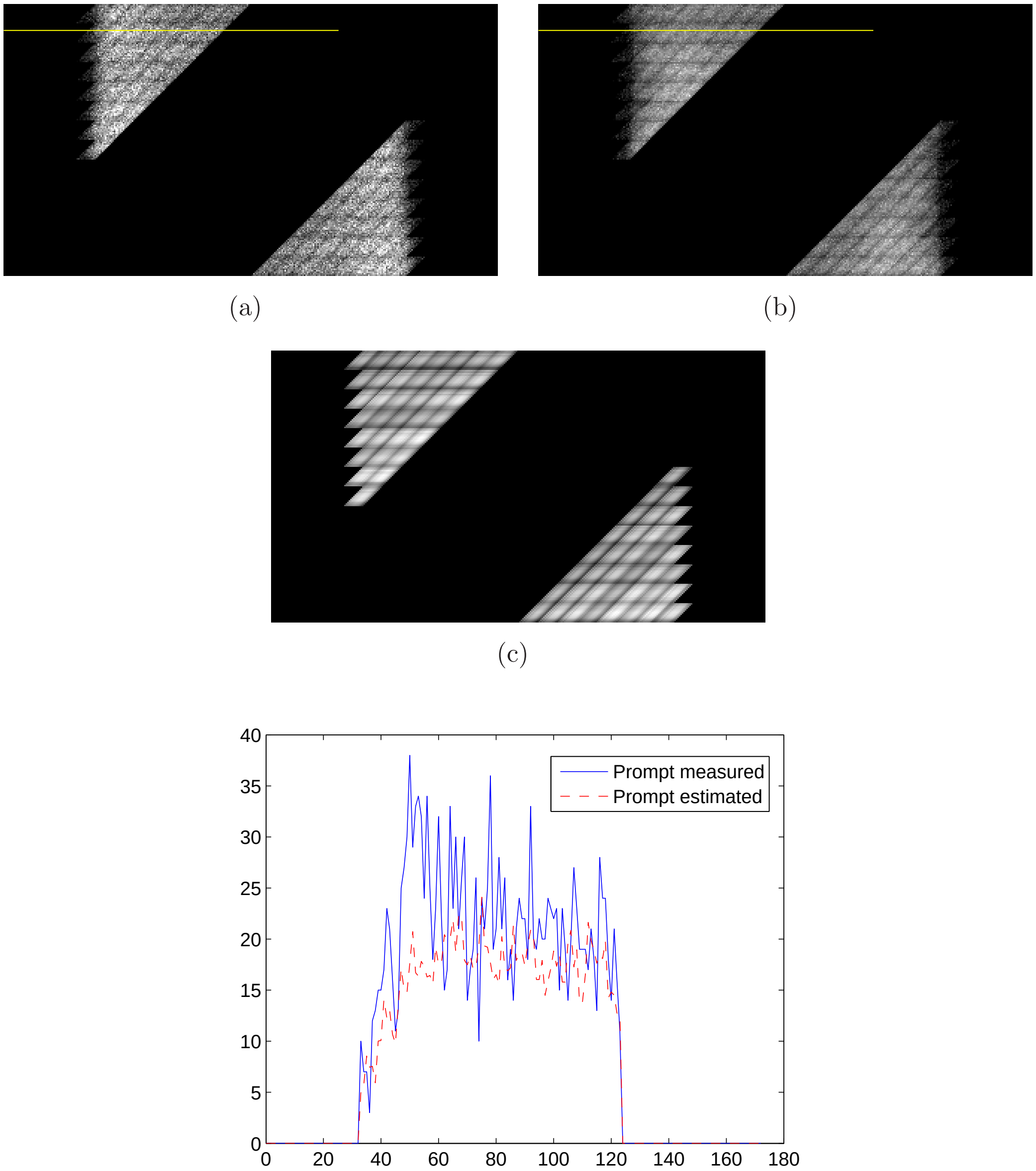


(a) (b)

(c)

(d)

Figure 6.18: IS sinograms and profile for the uniform cylinder acquisition for component-based normalization. (a) measured prompt sinogram; (b) estimated mean prompt sinogram; (c) ML normalization factor; (d) profile through the indicated row in sinograms (a) and (b). The image window is the same for (a) and (b).



(a) (b)

(c)

(d)

Figure 6.19: II sinograms and profile for the uniform cylinder acquisition for component-based normalization. (a) measured prompt sinogram; (b) estimated mean prompt sinogram; (c) ML normalization factor; (d) profile through the indicated row in sinograms (a) and (b). The image window is the same for (a) and (b).

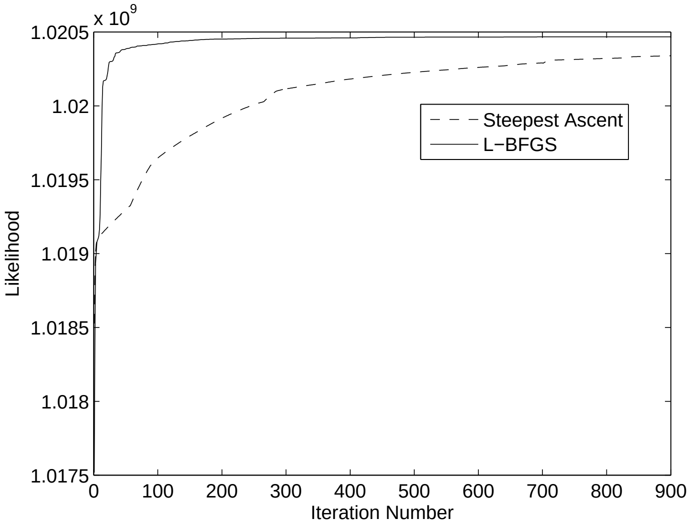


Figure 6.20: Comparison of convergence rate using steepest ascent versus L-BFGS for normalization factor estimation.

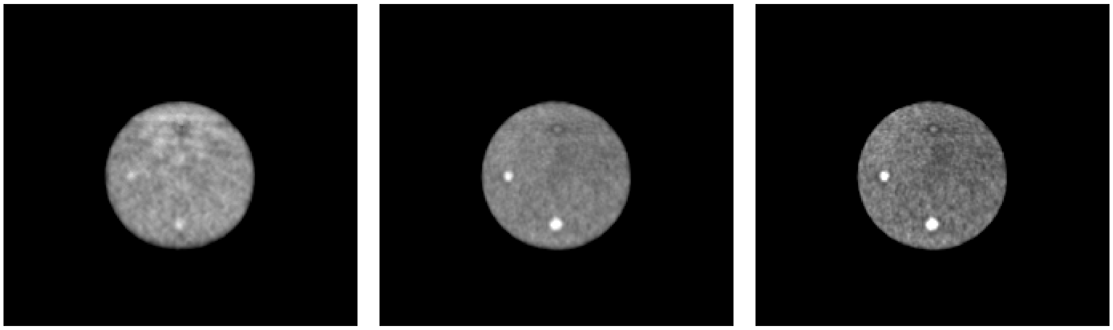
ML component-based normalization approach. Since the phantom was scanned twice (once without the insert, and once with the insert), the acquisition time was adjusted to account for the radioactive decay of ^{18}F; this ensured that approximately the same number of events occurred during each scan.

6.7 Discussion

In this chapter, we have developed and implemented a fully 3D reconstruction algorithm to jointly estimate an image using all three coincidence types. We used the geometrical factor derived in Chapter 5, but performed the calculations in 3D for all ring combinations. The body attenuation factor is also similar to that used in Chapter 5, except here it is calculated using a CT scan that has been converted to attenuation coefficients at 511 keV. A scatter estimate was calculated using the SSS method based on the Klein-Nishina formula. An initial scatter-free emission image estimate is required, as is an attenuation map of the body.

(a) (b)

Figure 6.21: Small tumor phantom images reconstructed using Siemens Biograph console. (a) CT slice that passes through the three spherical tumors; (b) PET reconstruction of the same slice using data from standalone PET scanner.



(a) (b) (c)

Figure 6.22: Small tumor phantom images reconstructed using our algorithm. The slice shown is the same as that in Figure 6.21. (a) Reconstruction using only SS data after 25 EM iterations; (b) reconstruction using II+IS+SS data after 25 EM iterations; (c) reconstruction using II+IS data after 20 EM iterations, initialized with the image in (b).

We used two rotating line sources for the direct normalization phantom. As shown in Figure 6.15, the measured data looks very similar to the forward-projected normalization phantom image. However, it is difficult to match the normalization phantom image exactly to the actual rotating line source trajectory. This creates a small mismatch between the two sinograms that is primarily visible when looking at the ratio sinogram. This shift leads to bias in the direct normalization factors, and is especially difficult to avoid when using thin line sources that are not perfectly aligned with the axial direction. This might not be as much of an issue if the rotation stage were built into the scanner, but in this work, the rotation stage is independent.

The bright vertical edges in the sinograms (which correspond to the radius at which each source rotates) are also difficult to model in the forward projection. Edges are particularly difficult to model because of insufficient sampling of the edge over the extent of the LOR, as already noted for the insert attenuation calculations in Section 6.4.

The Jaszczak phantom was reconstructed to demonstrate the performance of the direct normalization approach. As can be seen in the combined II+IS+SS reconstruction, there are some significant non-uniformity artifacts that are most likely due to the two problems mentioned above. The phantom is 21.6 cm diameter, which is barely smaller than the half-ring insert diameter. Therefore, part of the Jaszczak phantom projection is affected by the biased normalization factor in the region of the bright edge. Additionally, the direct normalization approach considered here did not estimate scatter, which could explain the bright region in the middle of the phantom reconstruction.

It was also determined after performing the rotating line source scan that the hot region within the line sources inadvertently covered slightly less than the entire axial FOV. In turn, this could have resulted in incorrect normalization factors for some of the sinograms, which would then be propagated to other slices in the fully 3D reconstruction through oblique sinograms. It seems unlikely that this would have caused the severe artifacts in the reconstruction, but it is a possibility.

To overcome these difficulties, the ML component-based method was developed and tested, as shown in Figure 6.22. In each fan-beam sinogram, the visible horizontal lines correspond to detector 1 of an LOR, and the visible diagonal lines correspond to detector 2 of the LOR. The product of the single crystal efficiencies therefore yields the patterns shown. When the insert is in the scanner FOV, the number of detected coincidences is significantly attenuated due to the insert material. We have modeled the insert attenuation and incorporated it as a fixed component of the component-based normalization method; it does not change from scan to scan. This allows us to obtain a separate CT scan of the body without the insert, and use that for the body attenuation factor.

The fact that the measured cylinder data and normalized mean cylinder data closely match each other implies that we have appropriately modeled the normalization factor components, as well as other system matrix factors. A profile through the SS and IS sinograms also reveals the similarity of the shape and scaling of the scatter and randoms tail. Some approximations are currently needed to determine the II scatter scaling factor due to the fact that no scatter tail is available for the scaling procedure. Additionally, the estimated mean profiles are noisy due to the lack of a variance reduction method when estimating the randoms. Such an approach may reduce the noise in what is intended to be a mean estimate of the randoms.

The small tumor phantom reconstructions demonstrate the ability of the insert system to resolve the smallest tumor. As mentioned above, this is a challenging object to image due to the very small size and low contrast ratio of the tumor. Since the details of the Siemens reconstruction algorithm were not available, it cannot be fully concluded from this study that the insert is able to resolve objects that the scanner itself cannot. To compare the performance of the insert system to a standalone system, the standalone scanner data should be reconstructed using the same algorithm implementation used for the insert system, but only using the scanner data. A separate set of normalization factors would need to be determined without the insert in the scanner FOV.

Nevertheless, the II+IS+SS reconstructions of this phantom are promising, and will lead naturally to further work in determining the optimal reconstruction strategy. For example, a more quantitative study would be useful to determine whether the best strategy is to perform a multi-stage reconstruction, where each stage is initialized by the preceding stage, e.g., first reconstruct using II+IS+SS data, then II+IS data, and finally, II data. Ordered subsets were not applied, so it may simply be the case that more iterations are needed to further enhance the image quality.

Much of the work described in this chapter was complicated by the fact that many factors depend on other factors being modeled accurately. We made use of standard methods where applicable, and modified them to suit the needs of the insert system as required. Since it is very challenging to separate out all the effects for independent study, there are many important directions for future research as a result of this work.

Chapter 7

Conclusions and Future Work

In this work, fully 3D statistical image reconstruction algorithms were developed and applied to helical CT and a novel half-ring PET insert system. While many of the challenges are specific to each modality, we have demonstrated that the same general approaches apply to both CT and PET. For both systems, we have modeled the scanners in their native 3D geometry to obtain the best possible image quality. Although fully 3D reconstruction is computationally expensive, we have developed algorithmic and hardware-based methods to optimize the reconstruction workflow as much as possible. As a result, we have been able to reconstruct clinically-sized datasets.

In each iteration of the statistical reconstruction algorithms for CT and PET, a comparison is performed between the measured data and the estimated data based on the current image iterate. If an accurate model for the measured data is used, the final reconstructed image will have little bias. Therefore, we have focused on developing accurate system models and integrating them into our reconstruction framework to obtain high-quality images.

The raw CT data that is output from the scanner has been preprocessed for effects like detector sensitivity variations, beam-hardening, and x-ray tube current modulation. Our efforts in modeling the CT data were therefore largely spent calculating the geometrical factor properly (and efficiently). In contrast, very little preprocessing has been applied to the raw PET data that is output from the scanner. Therefore, a significant effort was devoted to modeling and correcting for the various effects pertaining to the half-ring insert system.

7.1 CT-Specific Remarks

The CT work focused on maximizing the efficiency of the reconstruction algorithm by taking advantage of certain symmetries that underlie the complex multislice helical CT geometry. By formalizing the symmetries that exist, we were able to develop parallelized methods that perform well on large clinical datasets. In particular, we explicitly partitioned the data and image spaces across the processors in such a way that each processor performs exactly the same number of computations during the forward projections and backprojections, even when the OS method is employed. We showed that OS can increase the convergence rate significantly, and demonstrated that our convergence criterion was met in roughly 10-15 iterations with the use of a large number of subsets. Although this translates to about 30-45 times more computations than FBP-based techniques, it is still promising that the number of iterations required is relatively low.

The resolution-noise tradeoff results obtained using the Catphan phantom demonstrate the benefits of using statistical reconstruction compared to the analytical reconstruction methods we tested. By modeling the full extent of each voxel and detector element rather than using simpler voxel-driven backprojection methods, we ensure the resolution of each measurement is preserved. The clinical reconstructions appeared visually to be at least as good as the analytical methods considered, although further validation and analysis is needed.

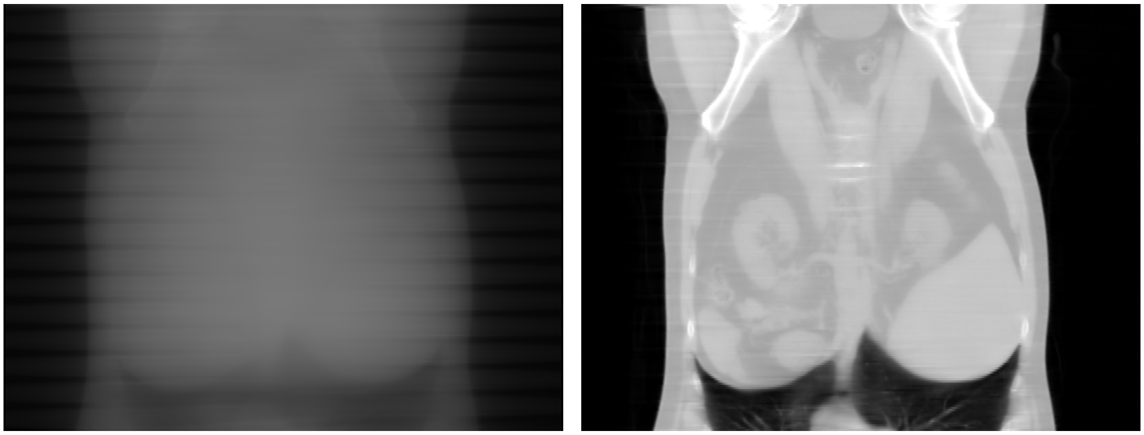
We also conducted a small study to apply two recently-developed geometry-independent methods to the 3D multislice helical CT image reconstruction problem. The first method reconstructs images without the missing data, and the second method seeks to jointly estimate the missing data and attenuation image in a statistical framework. We showed simulated data and clinical data reconstructions using these methods to handle the problem of transverse truncation. Along these lines, a method was proposed to handle the long object problem. We showed that it is possible to achieve accurate reconstructions within the FOV – and potentially outside the FOV as well – without using analytical data sufficiency or consistency conditions. One final advantage of these methods is that they provide a great deal of flexibility in their implementation for arbitrary system geometries. For example, this work could also apply to other 3D imaging modalities, such as C-arm cone-beam CT, in which the number of views is often limited. It could possibly even be used in 4D imaging (3D + time), in which each gated frame may have missing projections. It should be mentioned, however, that most of the missing data methods investigated in this work break the parallelism and symmetry constructs that exist when the complete data is available.

Regarding the regularization that was added to the AM algorithm, choosing suitable regularization parameters is notoriously difficult (especially in 3D). One could possibly test a range of parameters by performing several “trial” reconstructions on a downsampled dataset, or on just a few slices, and then attempt to scale the parameters accordingly for the full-scale problem. Other more systematic methods exist, but are also more computationally demanding.

There is some concern that the “quarter-detector offset” implemented on the commercial CT scanner we used for testing is not exactly a quarter-detector offset. Our reported results were based on an offset of -1.25 detectors in the data model instead of -0.25, due to the significantly better results. But even with an offset of -1.25 detectors, there is a very subtle rotating artifact that appears when scrolling through the stack of slices in the reconstructed image volume. If the actual detector offset can be known more accurately, there is a possibility of obtaining even higher-quality images, and of reducing the number of iterations required to reach convergence.

We will briefly discuss a second potential cause for the subtle rotating artifact, after describing this artifact in more detail. When visualized in a coronal or sagittal view, the image appears to have a striped pattern, as shown in Figure 7.1. These stripes are not horizontal, but are inclined slightly like the helical trajectory. As the iterations progress, the striped pattern becomes gradually less prevalent.

We believe the cause of this artifact relates to the inherently slow convergence of high frequencies using AM-type algorithms. Specifically, each voxel is seen a different number of times over the course of the scan; that is, there is nonuniform voxel illumination [92]. In analytical reconstruction algorithms (e.g., helical FDK), artifact-free backprojection requires that every voxel be seen the same number of times by each view angle. These algorithms account for this requirement with a weighting function that depends on the view angle and voxel. The AM algorithm does not need to use an explicit weighting function of this sort.



(a) (b)

Figure 7.1: Clinical reconstructions showing the striping artifact, and how it becomes less prevalent with more iterations. The AM algorithm was initialized with an image of all zeros, and OS was not used in the generation of these images. (a) Iteration 1; (b) iteration 50.

Starting from an initial uniform image of all zeros, the estimated backprojection in the first AM iterations will be mismatched from the measured data backprojection. Since the AM algorithm uses the ratio of these two backprojection images, the effects due to the nonuniform voxel sampling do not cancel out (i.e., since they are voxel and angle dependent, not just voxel dependent). This could lead to the striping pattern in early iterations. Assuming the system model used during the reconstruction is accurate, running more AM iterations should in theory lead to convergence, even when starting from an initial image of all zeros. However, even with OS, we have found that the stripes may take many iterations to remove completely. Therefore, this should be investigated further in future work to identify appropriate strategies for avoiding this artifact. For example, using a weighting function explicitly or initializing the image might help to reduce the discrepancy between the mean backprojection and measured data backprojection from the outset.

7.2 PET-Specific Remarks

The half-ring PET insert work described in this dissertation included MC simulations, experimental data acquisitions on our prototype system, data and system modeling, as well as 2D/3D reconstruction algorithm development and implementations. We placed emphasis on the system modeling aspects in this work, which include the geometrical, body attenuation, and normalization factors of the system matrix, along with estimation methods for randoms and scatter. Our reconstruction code allows the user to select any combination of coincidence types to be used in the reconstruction, which can be useful for evaluating the contributions of each coincidence type to the image.

We first introduced the 2D reconstruction framework, where the geometry, attenuation, and normalization were the only modeled components. Normalization was relatively straightforward since it provided a high-statistics dataset without scatter or randoms. We demonstrated that using the subvolume approach within the geometric factor calculations could lead to artifact-free reconstructions, even when activity was located above the side face of the insert crystals. We also showed in our reconstruction of 2D MC simulated data that DOI insert crystals could significantly reduce parallax error. However, the limited number of effects that we included in the MC simulation make it difficult to gain too much insight into the true performance of a real insert system.

For the real prototype system, we found that the LORs passing through long paths of LSO crystals cannot be modeled properly. We explained that the insert attenuation calculations fall apart for LORs passing tangentially through the insert crystals due to the invalid use of the linear approximation to the exponential function. For this reason, we have currently reduced the FOV to be inside the fan-width spanned by the insert. Further investigation on how to treat these LORs is warranted so that larger source distributions may be used in the future. It might be simplest to ignore these LORs completely, as they cannot be modeled properly and primarily contain randoms and scattered events from other parts of the FOV. Until this problem is addressed, the main application of the prototype insert system is limited to high-resolution head and neck imaging, as opposed to breast imaging.

Even when performing reconstructions on data in the reduced FOV, there are still some remaining issues that prevent uniform activity distributions from being reconstructed as uniformly as they should be. Most likely, these are caused by small discrepancies between the composite insert attenuation map and the true attenuation distribution of the insert. Addressing this could involve tweaking the attenuation coefficients slightly to obtain a better fit, as it appears that the reconstruction is very sensitive to the insert attenuation. A better mean randoms and/or scatter estimate may also help. Furthermore, poor alignment of the insert within the scanner may play some role in the uniformity problems as well; there are six degrees of freedom (horizontal, vertical, axial, yaw, pitch, and roll), but we only align the former three. Unfortunately, it is difficult to know where specifically the problem arises, as all the system matrix factors become interconnected in the reconstruction, and since the insert is cumbersome to align in a repeatable manner due to slight wobbling of the translation stage.

We did seem to have success using the ML estimation technique for component-based normalization, as seen by the good match between the measured and estimated uniform cylinder sinograms. The estimation technique is quite flexible, and can readily accommodate more components as necessary. For example, it might even be possible (and useful) to obtain a better estimate of the attenuation coefficients in a segmented insert attenuation map with this method. Additionally, ML estimation may be used to form a *mean* randoms estimate (rather than using the delayed counts directly as the estimate). Unlike whole-body clinical PET scanners, where there is usually a scatter-only tail outside the body projections, there is not necessarily a scatter-only tail in the II and IS measurements; whether it exists or not depends on the size of the object inside the insert FOV. Especially for the 20 cm diameter uniform normalization cylinder, it might therefore be useful to estimate the scatter scaling factors within the ML normalization method as well. This would only require that up to three more parameters be estimated (i.e., one each for II, IS, and SS). One other direction for the ML normalization framework is to try using source geometries other than a uniform cylinder, such as a uniform plane source or rotating rod sources. As long as an accurate normalization phantom image can be generated, the estimation procedure should be able to compute the component-based normalization factors.

Using the system model proposed in this work, it is possible to obtain high-resolution images that are jointly estimated from the three coincidence types. However, better resolution recovery may require the inclusion of a blur kernel in the system matrix to account for other sources of blurring, such as gamma ray acolinearity. It might also be interesting to compare the geometrical factor used in this work to a more accurate approach, such as an analytical detector response function model that does not subdivide the crystal volumes explicitly.

Other secondary future work for the half-ring PET insert reconstruction system would be the implementations of a regularization penalty and an ordered subsets method. Ordered subsets might be difficult to implement, however, since balancing the subsets in a non-circular ring is nontrivial. Alternative reconstruction algorithms that do not rely on ordered subsets for fast convergence, such as preconditioned conjugate gradient, could be investigated and compared to the EM algorithm.

7.3 Concluding Remarks

In this chapter, we highlighted our main contributions, and also showed that there is a large amount of interesting future work that could be pursued for both helical CT and the half-ring PET insert system. As the half-ring insert was the first clinical system ever built using the concept of virtual pinhole PET, it is to be expected that significant challenges would be encountered along the way. We have learned many valuable lessons from this work, and future virtual pinhole PET system designs should be able to make use of this newfound knowledge.

In conclusion, the work described in this dissertation has established a solid computational framework for statistical reconstruction, upon which improved techniques may be tested. It is anticipated that the fully 3D CT and PET reconstruction implementations and methodology presented here will be used in future projects.

Appendix A

Derivation of the Penalized AM Algorithm

To minimize (3.7), the two coupled variables in the argument of the potential function must be separated. The definition of a 1D convex function (instead of the more general Convex Decomposition Lemma applied in O'Sullivan and Benac [66]) can be used to achieve this. A function $f(t)$ is convex if

$$f[\alpha t_1 + (1-\alpha)t_2] \leq \alpha f(t_1) + (1-\alpha) f(t_2), \text{ where } 0 \leq \alpha \leq 1. \tag{A.1}$$

Using this property,

$$\begin{aligned}
\psi(\mu_j - \mu_{j'}) &= \psi\left\{\alpha\left[\frac{1}{\alpha}(\mu_j - \hat{\mu}_j) + (\hat{\mu}_j - \hat{\mu}_{j'})\right]\right. && \text{(A.2)} \\
&+ \left.(1-\alpha)\left[\frac{-1}{1-\alpha}(\mu_{j'} - \hat{\mu}_{j'}) + (\hat{\mu}_j - \hat{\mu}_{j'})\right]\right\} \\
&\leq \alpha\psi\left[\frac{1}{\alpha}(\mu_j - \hat{\mu}_j) + (\hat{\mu}_j - \hat{\mu}_{j'})\right] && \text{(A.3)} \\
&+ (1-\alpha)\psi\left[\frac{-1}{1-\alpha}(\mu_{j'} - \hat{\mu}_{j'}) + (\hat{\mu}_j - \hat{\mu}_{j'})\right].
\end{aligned}$$

To further simplify (A.3), let $\alpha \triangleq 1/2$ to obtain

$$\begin{aligned}
\psi(\mu_j - \mu_{j'}) &\leq \frac{1}{2}\psi\left[2(\mu_j - \hat{\mu}_j) + (\hat{\mu}_j - \hat{\mu}_{j'})\right] + \frac{1}{2}\psi\left[2(\hat{\mu}_{j'} - \mu_{j'}) + (\hat{\mu}_j - \hat{\mu}_{j'})\right] && \text{(A.4)} \\
&= \frac{1}{2}\left[\psi(2\mu_j - \hat{\mu}_j - \hat{\mu}_{j'}) + \psi(2\mu_{j'} - \hat{\mu}_j - \hat{\mu}_{j'})\right]. && \text{(A.5)}
\end{aligned}$$

We plug this surrogate for $\psi(\mu_j - \mu_{j'})$ into (3.8), and denote the modified penalty function by $\hat{R}(\boldsymbol{\mu})$. We now want to solve

$$\frac{\partial \Phi(\boldsymbol{\mu})}{\partial \mu_j} = \frac{\partial \mathcal{I}\left(\mathbf{d}||\mathbf{g}\right)}{\partial \mu_j} + \lambda \frac{\partial \hat{R}(\boldsymbol{\mu})}{\partial \mu_j} = 0. \tag{A.6}$$

In the final step before obtaining the voxel update function in the AM algorithm, the derivative of the I-divergence was determined to be

$$\frac{\partial \mathcal{I}\left(\mathbf{d}||\mathbf{g}\right)}{\partial \mu_j} = b_j - \hat{b}_j \exp\left[Z_j(\hat{\mu}_j - \mu_j)\right]. \tag{A.7}$$

Additionally, from (3.8) and (A.5),

$$\frac{\partial \hat{R}(\boldsymbol{\mu})}{\partial \mu_j} = \sum_{j' \in \mathcal{N}_j} w_{j,j'} \left. \frac{\partial \psi(t)}{\partial t} \right|_{t = 2\mu_j - \hat{\mu}_j - \hat{\mu}_{j'}}. \tag{A.8}$$

Finally, combining (A.6)-(A.8), we obtain the following equation for which we can readily use Newton's Method to solve for μ_j:

$$b_j - \hat{b}_j \exp\left[Z_j(\hat{\mu}_j - \mu_j)\right] + \lambda \sum_{j' \in \mathcal{N}_j} w_{j,j'} \left. \frac{\partial \psi(t)}{\partial t} \right|_{t = 2\mu_j - \hat{\mu}_j - \hat{\mu}_{j'}} = 0. \tag{A.9}$$

Appendix B

Helical CT Symmetry Validation

In this appendix, we mathematically prove the symmetry discussed in Chapter 3 for helical CT. In particular, we will evaluate $\tilde{h}(x, y, z|\beta, \gamma, \eta)$, the ideal point spread function, at multiples of a quarter rotation of the gantry.

To begin, we first express the x-ray source position, based on β, as

$$\begin{aligned} x_f &= R_f \cos\beta && \text{(B.1)} \\ y_f &= R_f \sin\beta && \text{(B.2)} \\ z_f &= \frac{\beta}{2\pi} z_{\text{feed}}. && \text{(B.3)} \end{aligned}$$

Using the geometry of the CT scanner shown in Figure 3.1, we can parametrically describe the ray originating at the source position and traveling in the direction of the detector (γ, η) by

$$\begin{aligned} x_{\text{ray}}(l) &= R_f \cos\beta - l\cos(\beta+\gamma) && \text{(B.4)} \\ y_{\text{ray}}(l) &= R_f \sin\beta - l\sin(\beta+\gamma) && \text{(B.5)} \\ z_{\text{ray}}(l) &= \frac{\beta}{2\pi} z_{\text{feed}} + l\tan\eta, && \text{(B.6)} \end{aligned}$$

where l is the ray parameter.

As discussed in Section 3.2.1, the ideal point spread function is nonzero only at points (x, y, z) along the x-ray path from source to detector. Therefore, we can write

$$\begin{aligned}
&\tilde{h}(x, y, z|\beta, \gamma, \eta) \triangleq \qquad \text{(B.7)}\\
&\qquad \delta[R_f \cos\beta - l\cos(\beta+\gamma) - x, R_f \sin\beta - l\sin(\beta+\gamma) - y, \frac{\beta}{2\pi} z_{\text{feed}} + l\tan\eta - z],
\end{aligned}$$

where $\delta(\cdot,\cdot,\cdot)$ is the Dirac delta function defined in three dimensions with the properties

$$\delta(\nu_1, \nu_2, \nu_3) \triangleq \begin{cases} 0, & \nu_1^2 + \nu_2^2 + \nu_3^2 \neq 0 \\ \infty, & \nu_1^2 + \nu_2^2 + \nu_3^2 = 0 \end{cases} \qquad \text{(B.8)}$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\nu_1, \nu_2, \nu_3)\, d\nu_1\, d\nu_2\, d\nu_3 \triangleq 1. \qquad \text{(B.9)}$$

Equation (B.7) is the ideal point spread function for any source-detector pair. We will show that the ideal point spread function for any source-detector pair beyond the first quarter rotation can be rewritten in terms of the corresponding base symmetry source-detector pair in the first quarter rotation (i.e., $0 \leq \beta < \pi/2$).

When the ray and point are both rotated by $2\pi k$ ($k \in \mathbb{Z}_{>0}$) and translated axially according to z_{feed}, the resulting point spread function is

$$\begin{aligned}
&\tilde{h}[x, y, z + k\cdot z_{\text{feed}}|\beta + 2\pi k, \gamma, \eta] = && \text{(B.10)}\\
&\qquad \delta\{R_f\cos(\beta+2\pi k) - l\cos(\beta+2\pi k+\gamma) - x, &&\\
&\qquad R_f\sin(\beta+2\pi k) - l\sin(\beta+2\pi k+\gamma) - y, &&\\
&\qquad \frac{\beta+2\pi k}{2\pi} z_{\text{feed}} + l\tan\eta - [z + k\cdot z_{\text{feed}}]\} &&\\
&\quad = \delta[R_f\cos\beta - l\cos(\beta+\gamma) - x, && \text{(B.11)}\\
&\qquad R_f\sin\beta - l\sin(\beta+\gamma) - y, &&\\
&\qquad \frac{\beta}{2\pi} z_{\text{feed}} + l\tan\eta - z] &&\\
&\quad = \tilde{h}(x, y, z|\beta, \gamma, \eta). && \text{(B.12)}
\end{aligned}$$

When the ray and point are both rotated by $\pi/2 + 2\pi k$ ($k \in \mathbb{Z}_{\geq 0}$) and translated axially according to z_{feed}, the resulting point spread function is

$$\begin{aligned}
&\tilde{h}[-y, x, z + z_{\text{feed}}(1/4 + k)|\beta + \pi/2 + 2\pi k, \gamma, \eta] = \qquad \text{(B.13)}\\
&\qquad \delta\{R_f \cos(\beta + \pi/2 + 2\pi k) - l\cos(\beta + \pi/2 + 2\pi k + \gamma) + y,\\
&\qquad R_f \sin(\beta + \pi/2 + 2\pi k) - l\sin(\beta + \pi/2 + 2\pi k + \gamma) - x,\\
&\qquad \frac{\beta + \pi/2 + 2\pi k}{2\pi} z_{\text{feed}} + l\tan\eta - [z + z_{\text{feed}}(1/4 + k)]\}\\
&\quad = \delta[-R_f \sin\beta + l\sin(\beta + \gamma) + y, \qquad \text{(B.14)}\\
&\qquad R_f \cos\beta - l\cos(\beta + \gamma) - x,\\
&\qquad \frac{\beta}{2\pi} z_{\text{feed}} + l\tan\eta - z]\\
&\quad = \tilde{h}(x, y, z|\beta, \gamma, \eta). \qquad \text{(B.15)}
\end{aligned}$$

When the ray and point are both rotated by $\pi + 2\pi k$ ($k \in \mathbb{Z}_{\geq 0}$) and translated axially according to z_{feed}, the resulting point spread function is

$$\begin{aligned}
&\tilde{h}[-x, -y, z + z_{\text{feed}}(1/2 + k)|\beta + \pi + 2\pi k, \gamma, \eta] = \qquad \text{(B.16)}\\
&\qquad \delta\{-R_f \cos(\beta + \pi + 2\pi k) + l\cos(\beta + \pi + 2\pi k + \gamma) + x,\\
&\qquad -R_f \sin(\beta + \pi + 2\pi k) + l\sin(\beta + \pi + 2\pi k + \gamma) + y,\\
&\qquad \frac{\beta + \pi + 2\pi k}{2\pi} z_{\text{feed}} + l\tan\eta - [z + z_{\text{feed}}(1/2 + k)]\}\\
&\quad = \delta[-R_f \cos\beta + l\cos(\beta + \gamma) + x, \qquad \text{(B.17)}\\
&\qquad -R_f \sin\beta + l\sin(\beta + \gamma) + y,\\
&\qquad \frac{\beta}{2\pi} z_{\text{feed}} + l\tan\eta - z]\\
&\quad = \tilde{h}(x, y, z|\beta, \gamma, \eta). \qquad \text{(B.18)}
\end{aligned}$$

When the ray and point are both rotated by $3\pi/2 + 2\pi k$ ($k \in \mathbb{Z}_{\geq 0}$) and translated axially according to z_{feed}, the resulting point spread function is

$$
\begin{aligned}
&\tilde{h}[y, -x, z + z_{\text{feed}}(3/4 + k)|\beta + 3\pi/2 + 2\pi k, \gamma, \eta] = \qquad \text{(B.19)}\\
&\quad \delta\{R_f \cos(\beta + 3\pi/2 + 2\pi k) - l \cos(\beta + 3\pi/2 + 2\pi k + \gamma) - y,\\
&\quad R_f \sin(\beta + 3\pi/2 + 2\pi k) - l \sin(\beta + 3\pi/2 + 2\pi k + \gamma) + x,\\
&\quad \frac{\beta + 3\pi/2 + 2\pi k}{2\pi} z_{\text{feed}} + l \tan\eta - [z + z_{\text{feed}}(3/4 + k)]\}\\
&= \delta[R_f \sin\beta - l \sin(\beta + \gamma) - y, \qquad \text{(B.20)}\\
&\quad -R_f \cos\beta + l \cos(\beta + \gamma) + x,\\
&\quad \frac{\beta}{2\pi} z_{\text{feed}} + l \tan\eta - z]\\
&= \tilde{h}(x, y, z|\beta, \gamma, \eta). \qquad \text{(B.21)}
\end{aligned}
$$

Finally, when there is no quarter detector offset, axial mirror symmetry can also be obtained. This symmetry, if available, eliminates the need to compute and store the system matrix for all detector rows. Only half the rows need to be computed and stored, as the other half follows symmetrically.

Consider the ray that has opposite source, fan, and cone angles from the base source-detector pair. Also consider the point that is the reflection of the base point across the $y = 0$ and $z = 0$ planes. Note that although this point and ray are located in front of the axial FOV of the scanner, they can be rotated and translated axially into the FOV using the above quarter rotation symmetry. Mathematically, the axial mirror symmetry follows as

$$\begin{aligned}
&\tilde{h}(x,-y,-z|-\beta,-\gamma,-\eta) = && \text{(B.22)}\\
&\qquad \delta[R_f\cos(-\beta) - l\cos(-\beta-\gamma) - x, &&\\
&\qquad R_f\sin(-\beta) - l\sin(-\beta-\gamma) + y, &&\\
&\qquad \frac{-\beta}{2\pi} z_{\text{feed}} + l\tan(-\eta) + z] &&\\
&= \delta[R_f\cos\beta - l\cos(\beta+\gamma) - x, && \text{(B.23)}\\
&\qquad -R_f\sin\beta + l\sin(\beta+\gamma) + y, &&\\
&\qquad \frac{-\beta}{2\pi} z_{\text{feed}} - l\tan\eta + z] &&\\
&= \tilde{h}(x,y,z|\beta,\gamma,\eta). && \text{(B.24)}
\end{aligned}$$

In this appendix, we have focused our discussion on the symmetry that is present in $\tilde{h}(x,y,z|\beta,\gamma,\eta)$. As explained next, the integrations over the extent of each voxel and detector element in (3.10) and (3.11), respectively, do not change the symmetry. For the integral over each voxel, the ideal point spread function is simply being evaluated at different points. For the integral over the detector element surface, the symmetry applies to each infinitesimally thin subray that lies within the integration bounds. The $\cos\eta$ factor in (3.11) does not present any difficulty for the axial mirror symmetry since $\cos(-\eta) = \cos\eta$.

References

[1] S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero. Convergent incremental optimization transfer algorithms: application to tomography. *IEEE Trans. Med. Imag.*, 25(3):283–296, 2006.

[2] T. Akenine-Möller and E. Haines. *Real-Time Rendering.* AK Peters, Ltd, 2nd edition, 2002.

[3] J. Amanatides and A. Woo. A fast voxel traversal algorithm for ray tracing. In *Eurographics*, 1987.

[4] R. D. Badawi, M. A. Lodge, and P. K. Marsden. Algorithms for calculating detector efficiency normalization coefficients for true coincidences in 3D PET. *Phys. Med. Biol.*, 43:189–205, 1998.

[5] B. Bai, Q. Li, C. H. Holdsworth, E. Asma, Y. C. Tai, A. Chatziioannou, and R. M. Leahy. Model-based normalization for iterative 3D PET image reconstruction. *Phys. Med. Biol.*, 47:2773–2784, 2002.

[6] D. L. Bailey, D. W. Townsend, P. E. Kinahan, S. Grootoonk, and T. Jones. An investigation of factors affecting detector and geometric correction in normalization of 3-D PET data. *IEEE Trans. Nucl. Sci.*, 43(6):3300–3307, 1996.

[7] S. Basu and Y. Bresler. $O(N^2 \log_2 N)$ filtered backprojection reconstruction algorithm for tomography. *IEEE Trans. Imag. Proc.*, 9(10):1760–1773, October 2000.

[8] S. Basu and B. De Man. Branchless distance driven projection and backprojection. In C. A. Bouman, E. L. Miller, and I. Pollak, editors, *Proc. SPIE: Electronic Imaging*, volume 6065, 2006.

[9] T. M. Benson and J. Gregor. Framework for iterative cone-beam micro-CT reconstruction. *IEEE Trans. Nucl. Sci.*, 52(5):1335–1340, 2005.

[10] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. Imag. Proc.*, 5(3):480–492, 1996.

[11] C. Burger, G. Goerres, S. Schoenes, A. Buck, A. H. R. Lonn, and G. K. von Schulthess. PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients. *Eur. J. Nucl. Med.*, 29(7):922–927, 2002.

[12] M. E. Casey, H. Gadagkar, and D. Newport. A component based method for normalization in volume PET. In *Proc. 3rd Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 67–71, 1995.

[13] A. Chatziioannou. VP-PET: A new imaging modality? *J. Nucl. Med.*, 49(3): 345–346, 2008.

[14] S. Cho, Q. Li, S. Ahn, B. Bai, and R. M. Leahy. Iterative image reconstruction ising inverse Fourier rebinning for fully 3-D PET. *IEEE Trans. Med. Imag.*, 26 (5):745–756, 2007.

[15] P. E. Danielsson, P. Edholm, J. Eriksson, and S. M. Magnusson. Towards exact reconstruction for helical cone-beam scanning of long objects. In D W Townsend and P E Kinahan, editors, *Proc. 1997 Meeting on Fully 3D Reconstruction in Radiology and Nuclear Medicine*, pages 141–144, Pittsburgh, PA, 1997.

[16] M. E. Daube-Witherspoon and G. Muehllehner. Treatment of axial data in three-dimensional PET. *J. Nucl. Med.*, 28:1717–1724, 1987.

[17] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49:2463–2475, 2004.

[18] M. Defrise, P. E. Kinahan, D. W. Townsend, C. Michel, M. Sibomana, and D. F. Newport. Exact and approximate rebinning algorithms for 3-D PET data. *IEEE Trans. Med. Imag.*, 16(2):145–158, 1997.

[19] M. Defrise, F. Noo, and H. Kudo. A solution to the long-object problem in helical cone-beam tomography. *Phys. Med. Biol.*, 45:623–643, 2000.

[20] M. Defrise, F. Noo, R. Clackdoyle, and H. Kudo. Truncated Hilbert transform and image reconstruction from limited tomographic data. *Inverse Problems*, 22: 1037–1053, 2006.

[21] I. A. Elbakri and J. A. Fessler. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans. Med. Imag.*, 21(2):89–99, 2002.

[22] H. Erdoğan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44:2835–2851, 1999.

[23] F. H. Fahey. Data acquisition in PET imaging. *J. Nucl. Med. Tech.*, 30(2): 39–49, 2002.

[24] J. A. Fessler. Statistical image reconstruction methods for transmission tomography. In M. Sonka and J. M. Fitzpatrick, editors, *Handbook of Medical Imaging: Medical Image Processing and Analysis*, volume 2, pages 1–70, Bellingham, WA, 2000. SPIE.

[25] J. A. Fessler and A. O. Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Trans. Imag. Proc.*, 4(10):1417–1429, 1995.

[26] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. *IEEE Trans. Imag. Proc.*, 5(9):1346–1358, 1996.

[27] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Trans. Med. Imag.*, 16(2):166–175, 1997.

[28] R. Freifelder and J. S. Karp. Dedicated PET scanners for breast imaging. *Phys. Med. Biol.*, 42:2463–2480, 1997.

[29] L. Fu, J. Liao, and J. Qi. Evaluation of 2D ROI image reconstruction using ML-EM method from truncated projections. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 2236–2241, 2006.

[30] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imag.*, 9:84–93, 1990.

[31] D. Hogg, K. Thielemans, T. Spinks, and N. Spyrou. Maximum-likelihood estimation of normalisation factors for PET. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 2065–2069, 2001.

[32] D. Hogg, K. Thielemans, S. Mustafovic, and T. J. Spinks. A study of bias for various iterative reconstruction methods in PET. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 1519–1523, 2002.

[33] J. Hsieh. Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise. *Med. Phys.*, 25(11):2139–2147, 1998.

[34] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash, and T. J. Myers. A novel reconstruction algorithm to extend the CT scan field-of-view. *Med. Phys.*, 31:2385–2391, 2004.

[35] J. Hu, J. Qi, J. S. Huber, W. W. Moses, and R. H. Huesman. MAP image reconstruction for arbitrary geometry PET systems with application to a prostate-specific scanner. In *The 8th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2005.

[36] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 13(4):601–609, 1994.

[37] R. H. Huesman, G. J. Klein, W. W. Moses, J. Qi, B. W. Reutter, and P. R. G. Virador. List-mode maximum-likelihood reconstruction applied to positron emission mammography (PEM) with irregular sampling. *IEEE Trans. Med. Imag.*, 19:532–537, 2000.

[38] A. A. Isola, A. Ziegler, T. Koehler, W. J. Niessen, and M. Grass. Motion-compensated iterative cone-beam CT image reconstruction with adapted blobs as basis functions. *Phys. Med. Biol.*, 53:6777–6797, 2008.

[39] C. A. Johnson and A. Sofer. A data-parallel algorithm for tomographic image reconstruction. In *The 7th Symposium on the Frontiers of Massively Parallel Computation*, pages 126–137, Washington D.C., 1999. IEEE Computer Society Press.

[40] M. D. Jones and R. Yao. Parallel programming for OSEM reconstruction with MPI, OpenMP, and hybrid MPI-OpenMP. In *IEEE Nuclear Science Symposium Conference Record*, pages 3036–3042, 2004.

[41] M. Kachelriess, M. Knaup, and O. Bockenbach. Hyperfast parallel-beam and cone-beam backprojection using the cell general purpose hardware. *Med. Phys.*, 34(4):1474–1486, Apr 2007.

[42] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging.* IEEE Press, New York, NY, 1988.

[43] M. K. Kalra, M. M. Maher, T. L. Toth, L. M. Hamberg, M. A. Blake, J.-A. Shepard, and S. Saini. Strategies for CT radiation dose optimization. *Radiology*, 230:619–628, 2004.

[44] A. Katsevich. An improved exact filtered backprojection algorithm for spiral computed tomography. *Advances in Applied Mathematics*, 32(4):681–697, 2003.

[45] P. E. Kinahan, B. H. Hasegawa, and T. Beyer. X-ray-based attenuation correction for positron emission tomography/computed tomography scanners. *Sem. Nucl. Med.*, 33:166–179, 2003.

[46] M. Knaup, W. A. Kalender, and M. Kachelriess. Statistical cone-beam CT image reconstruction using the cell broadband engine. In *IEEE Nuclear Science Symposium Conference Record*, pages 2837–2840, 2006.

[47] J. S. Kole and F. J. Beekman. Parallel statistical image reconstruction for cone-beam x-ray CT on a shared memory computation platform. *Phys. Med. Biol.*, 50:1265–1272, 2005.

[48] J. S. Kole and F. J. Beekman. Evaluation of accelerated iterative x-ray CT image reconstruction using floating point graphics hardware. *Phys. Med. Biol.*, 51(4):875–889, Feb 2006.

[49] Hiroyuki Kudo, Frédéric Noo, and Michel Defrise. Quasi-exact filtered backprojection algorithm for long-object problem in helical cone-beam tomography. *IEEE Trans. Med. Imag.*, 19(9):902–921, 2000.

[50] P. J. La Rivière. Monotonic iterative reconstruction algorithms for targeted reconstruction in emission and transmission computed tomography. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 2924–2928, 2006.

[51] K. Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Trans. Med. Imag.*, 9(4):439–446, 1990.

[52] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *J. Computer Assisted Tomography*, 8(2):306–316, 1984.

[53] G. M. Lasio, B. R. Whiting, and J. F. Williamson. Statistical reconstruction for x-ray computed tomography using energy-integrating detectors. *Phys. Med. Biol.*, 52:2247–2266, 2007.

[54] T. K. Lewellen. Recent developments in PET detector technology. *Phys. Med. Biol.*, 53:R287–R317, 2008.

[55] P. J. Markiewicz, A. J. Reader, M. Tamal, P. J. Julyan, and D. L. Hastings. An advanced analytic method incorporating the geometrical properties of scanner and radiation emissions into the system model for the true component of 3D PET data. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, 2005.

[56] P. J. Markiewicz, M. Tamal, P. J. Julyan, D. L. Hastings, and A. J. Reader. High accuracy multiple scatter modelling for 3D whole body PET. *Phys. Med. Biol.*, 52:829–847, 2007.

[57] S. Matej and R. M. Lewitt. Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Trans. Med. Imag.*, 15(1):68–78, 1996.

[58] S. R. Meikle and R. D. Badawi. Quantitative techniques in PET. In D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, editors, *Positron Emission Tomography: Basic Sciences*. Springer, 2005.

[59] S. Moehrs, M. Defrise, N. Belcari, A. Del Guerra, A. Bartoli, S. Fabbri, and G. Zanetti. Multi-ray-based system matrix generation for 3D PET reconstruction. *Phys. Med. Biol.*, 53:6925–6945, 2008.

[60] C. Mora and M. Rafecas. Polar pixels for high resolution small animal PET. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 2812–2817, 2006.

[61] E. Ü. Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. *IEEE Trans. Med. Imag.*, 13(4):687–701, 1994.

[62] E. Ü. Mumcuoğlu, R. M. Leahy, and S. R. Cherry. Bayesian reconstruction of PET images: methodology and performance analysis. *Phys. Med. Biol.*, 41: 1777–1807, 1996.

[63] F. Noo, J. Pack, and D. Heuscher. Exact helical reconstruction using native cone-beam geometries. *Phys. Med. Biol.*, 48:3787–3818, 2003.

[64] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans. Iterative reconstruction for helical CT: a simulation study. *Phys. Med. Biol.*, 43(4):729–737, Apr 1998.

[65] Naoaki Okazaki. libLBFGS: a library of limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS). Obtained from 'http://www.chokkan.org/software/liblbfgs/', 2009.

[66] J. A. O'Sullivan and J. Benac. Alternating minimization algorithms for transmission tomography. *IEEE Trans. Med. Imag.*, 26(3):283–297, Mar 2007.

[67] D. Pal. *Image reconstruction for a high resolution PET-insert scanner*. PhD thesis, Washington Univ. in St. Louis, St. Louis, MO, 2008.

[68] D. Pal, J. A. O'Sullivan, H. Wu, and Y.-C. Tai. Generalized 3D kernel computation method and its application in PET insert system. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 1711–1714, 2006.

[69] D. Pal, D. B. Keesing, J. A. O'Sullivan, S. Komarov, and Y.-C. Tai. Image reconstruction algorithm for a special geometry of the PET-insert system. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 2791–2798, 2007.

[70] D. Pal, J. A. O'Sullivan, H. Wu, M. Janecek, and Y.-C. Tai. 2D linear and iterative reconstruction algorithms for a PET-insert scanner. *Phys. Med. Biol.*, 52:4293–4310, 2007.

[71] V. Y. Panin, F. Kehren, C. Michel, and M. Casey. Fully 3-D PET reconstruction with system matrix derived from point source measurements. *IEEE Trans. Med. Imag.*, 25(7):907–921, 2006.

[72] D. G. Politte and B. R. Whiting. A fast, accurate projector/backprojector pair for use with iterative image reconstruction for two-dimensional fan-beam transmission imaging. Technical report, Washington University in St. Louis, Electronic Radiology Laboratory, St. Louis, Missouri, November 2004.

[73] D. G. Politte, S. Yan, J. A. O'Sullivan, D. L. Snyder, and B. R. Whiting. Implementation of alternating minimization algorithms for fully 3D CT imaging. In C A Bouman and E L Miller, editors, *Proc. SPIE: Computational Imaging III*, volume 5674, pages 362–373, 2005.

[74] J. Qi and R. M. Leahy. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *IEEE Trans. Med. Imag.*, 19(5):493–506, 2000.

[75] J. Qi and R. M. Leahy. Iterative reconstruction techniques in emission computed tomography. *Phys. Med. Biol.*, 51:R541–R578, 2006.

[76] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar. High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner. *Phys. Med. Biol.*, 43:1001–1013, 1998.

[77] M. Rafecas, B. Mosler, M. Dietz, M. Pögl, A. Stamatakis, D. P. McElroy, and S. I. Ziegler. Use of a Monte Carlo-based probability matrix for 3-D iterative reconstruction of MADPET-II data. *IEEE Trans. Nucl. Sci.*, 51(5):2597–2605, 2004.

[78] N. Rehfeld and M. Alber. The influence of noise in full Monte Carlo ML-EM and dual matrix reconstructions in positron emission tomography. *Med. Phys.*, 33(9):3498–3507, 2006.

[79] K. J. Ruchala, G. H. Olivera, J. M. Kapatoes, P. J. Reckwerdt, and T. R. Mackie. Methods for improving limited field-of-view radiotherapy reconstructions using imperfect *a priori* images. *Med. Phys.*, 29:2590–2605, 2002.

[80] S. Schaller, T. Flohr, K. Klingenbeck, J. Krause, T. Fuchs, and W. A. Kalender. Spiral interpolation algorithm for multislice spiral CT – part I: Theory. *IEEE Trans. Med. Imag.*, 19(9):822–834, 2000.

[81] S. Schaller, F. Noo, F. Sauer, K. C. Tam, G. Lauritsch, and T. Flohr. Exact radon rebinning algorithm for the long object problem in helical cone-beam CT. *IEEE Trans. Med. Imag.*, 19:361–375, 2000.

[82] S. Schaller, K. Stierstorfer, H. Bruder, M. Kachelriess, and T. Flohr. Novel approximate approach for high-quality image reconstruction in helical cone beam CT at arbitrary pitch. In *Proc. SPIE: Medical Imaging*, volume 4322, pages 113–127, 2001.

[83] S. Schaller, J. E. Wildberger, R. Raupach, M. Niethammer, K. Klingenbeck-Regn, and T. Flohr. Spatial domain filtering for fast modification of the tradeoff between image sharpness and pixel noise in computed tomography. *IEEE Trans. Med. Imag.*, 22(7):846–853, 2003.

[84] M. Schellmann, T. Kosters, and S. Gorlatch. Parallelization and runtime prediction of the listmode OSEM algorithm for 3D PET reconstruction. In *IEEE Nuclear Science Symposium Conference Record*, pages 2190–2195, 2006.

[85] W. P. Segars. *Development of a new dynamic NURBS-based cardiac-torso (NCAT) phantom*. PhD thesis, The University of North Carolina, May 2001.

[86] K. Sheng, R. Jeraj, R. Shaw, T. R. Mackie, and B. Paliwal. Imaging dose management using multi-resolution in CT-guided radiation therapy. *Phys. Med. Biol.*, 50:1205–1219, 2005.

[87] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, 1(2):113–121, 1982.

[88] E. Y. Sidky and X. Pan. Accurate image reconstruction in circular cone-beam computed tomography by total variation minimization: a preliminary investigation. In *IEEE Nuclear Science Symposium/Medical Imaging Conference*, pages 2904–2907, 2006.

[89] D. L. Snyder, J. A. O'Sullivan, R. J. Murphy, D. G. Politte, B. R. Whiting, and J. F. Williamson. Image reconstruction for transmission tomography when projection data are incomplete. *Phys. Med. Biol.*, 51:5603–5619, 2006.

[90] K. Sourbelle, M. Kachelriess, and W. A. Kalender. Reconstruction from truncated projections in CT using adaptive detruncation. *Eur. Radiol.*, 15:1008–1014, 2005.

[91] J. W. Stayman and J. A. Fessler. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Trans. Med. Imag.*, 19(6):601–615, 2000.

[92] S. Steckmann, M. Knaup, and M. Kachelriess. High performance cone-beam spiral backprojection with voxel-specific weighting. *Phys. Med. Biol.*, 54:3691–3708, 2009.

[93] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr. Weighted FBP – a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch. *Phys. Med. Biol.*, 49:2209–2218, 2004.

[94] D. Strul, R. B. Slates, M. Dahlbom, S. R. Cherry, and P. K. Marsden. An improved analytical detector response function model for multilayer small-diameter PET scanners. *Phys. Med. Biol.*, 48:979–994, 2003.

[95] Y.-C. Tai, H. Wu, and M. Janecek. Initial study of an asymmetric PET system dedicated to breast cancer imaging. *IEEE Trans. Nucl. Sci.*, 53:121–126, 2006.

[96] Y.-C. Tai, H. Wu, D. Pal, and J. A. O'Sullivan. Virtual-pinhole PET. *J. Nucl. Med.*, 49:471–479, 2008.

[97] X. Tang, J. Hsieh, R. A. Nilsen, S. Dutta, D. Samsonov, and A. Hagiwara. A three-dimensional-weighted cone beam filtered backprojection (CB-FBP) algorithm for image reconstruction in volumetric CT – helical scanning. *Phys. Med. Biol.*, 51:855–874, 2006.

[98] C. Terboven, D. an Mey, and S. Sarholz. Openmp on multicore architectures. In B. Chapman et al., editor, *Lecture Notes in Computer Science*, volume 4935, chapter IWOMP 2007, pages 54–64. Springer, 2008.

[99] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multislice helical CT. *Med. Phys.*, 34(11):4526–4544, 2007.

[100] M. Tianyu, Z. Rong, and J. Yongjie. Communication optimization and auto load balancing in parallel OSEM algorithm for fully-3D SPECT reconstruction. In *IEEE Nuclear Science Symposium Conference Record*, pages 2695–2699, 2005.

[101] H. L. van Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley & Sons, Inc., 2001.

[102] G. Wang and H. Yu. An outlook on x-ray CT research and development. *Med. Phys.*, 35(3):1051–1064, 2008.

[103] C. C. Watson. New, faster, image-based scatter correction for 3D PET. *IEEE Trans. Nucl. Sci.*, 47(4):1587–1594, 2000.

[104] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson. Properties of preprocessed sinogram data in x-ray computed tomography. *Med. Phys.*, 33(9):3290–3303, 2006.

[105] J. F. Williamson, B. R. Whiting, J. Benac, R. J. Murphy, G. J. Blaine, J. A. O'Sullivan, D. G. Politte, and D. L. Snyder. Prospects for quantitative computed tomography imaging in the presence of foreign metal bodies using statistical image reconstruction. *Med. Phys.*, 29(10):2404–2418, 2002.

[106] H. Wu, T. Y. Song, D. Pal, D. B. Keesing, S. Komarov, J. A. O'Sullivan, and Y.-C. Tai. A high resolution PET insert system for clinical PET/CT scanners. In *IEEE Nuclear Science Symposium Conference Record*, pages 5442–5444, 2008.

[107] F. Xu and K. Mueller. Real-time 3D computed tomographic reconstruction using commodity graphics hardware. *Phys. Med. Biol.*, 52:3405–3419, 2007.

[108] T. Yamaya, N. Hagiwara, T. Obi, B. Yamaguchi, N. Ohymama, K. Kitamura, T. Haseqawa, H. Haneishi, E. Yoshida, N. Inadama, and H. Murayama. Transaxial system models for jPET-D4 image reconstruction. *Phys. Med. Biol.*, 50:5339–5355, 2005.

[109] W. Yao and K. Leszczynski. Analytically derived weighting factors for transmission tomography cone beam projections. *Phys. Med. Biol.*, 54:513–533, 2009.

[110] M. Yavuz and J. A. Fessler. Statistical image reconstruction methods for randoms-precorrected PET scans. *Med. Image Anal.*, 2:369–378, 1998.

[111] G. L. Zeng, G. T. Gullberg, P. E. Christian, and D. Gagnon. Cone-beam iterative reconstruction of a segment of a long object. *IEEE Trans. Nucl. Sci.*, 49:37–41, 2002.

[112] B. Zhang and G. Zeng. Comparison study of iterative reconstruction from truncated projections in 2D and 3D imaging. *J. Nucl. Med.*, 47S1:122P, 2006.

[113] A. Ziegler, T. Nielsen, and M. Grass. Iterative reconstruction of a region of interest for transmission tomography. *Med. Phys.*, 35(4):1317–1327, 2008.

[114] Y. Zou and X. Pan. Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT. *Phys. Med. Biol.*, 49:941–959, 2004.

Vita

Daniel Brian Keesing

Date of Birth	September 27, 1981
Place of Birth	Stamford, Connecticut
Degrees	B.S. Electrical Engineering, Tufts University, 2004 M.S. Biomedical Engineering, Washington University, 2006 Ph.D. Biomedical Engineering, Washington University, 2009
Professional and Honor Societies	Institute of Electrical and Electronics Engineers (IEEE) Society for Industrial and Applied Mathematics (SIAM) Tau Beta Pi Honor Society Eta Kappa Nu Honor Society
Selected Publications	**D. Keesing**, J. O'Sullivan, H. Wu, S. Komarov, Y.-C. Tai. Fully 3D Statistical Image Reconstruction Algorithm for Half-Ring PET Insert System, in preparation. **D. Keesing**, D. Politte, B. Whiting, J. O'Sullivan. Efficient Implementation of a Fully 3D Statistical Image Reconstruction Algorithm for Multislice Helical CT, in preparation. **D. Keesing**, D. Pal, J. O'Sullivan, S. Komarov, Y.-C. Tai. System Modeling of a DOI-Capable PET Insert Device for Breast Imaging, *IEEE Medical Imaging Conference*, 2008. H. Wu, T. Y. Song, D. Pal, **D. Keesing**, S. Komarov, J. O'Sullivan, Y.-C. Tai. A High Resolution PET Insert System for Clinical PET/CT Scanners, *IEEE Medical Imaging Conference*, 2008. D. Pal, **D. Keesing**, J. O'Sullivan, S. Komarov, Y.-C. Tai. Image Reconstruction Algorithm for a Special Geometry of

the PET-Insert System, *IEEE Medical Imaging Conference*, 2007.

D. Keesing, J. O'Sullivan, D. Politte, B. Whiting, D. Snyder. Missing Data Estimation for Fully 3D Spiral CT Image Reconstruction, *SPIE Medical Imaging*, 2007.

D. Keesing, J. O'Sullivan, D. Politte, B. Whiting. Parallelization of a Fully 3D CT Iterative Reconstruction Algorithm, *IEEE International Symposium on Biomedical Imaging*, 2006.

August 2009